Received 19 March; accepted 18 June 1984.

1. Bate, C. M. & Grunewald, E. B. *J. Embryol. exp. Morph.* 61, 317–330 (1981).
2. Goodman, C. S. & Bate, M. *Trends Neurosci.* 4, 163–169 (1981).
3. Goodman, C. S., Raper, J. A., Ho, R. K. & Chang, S. *Symp. Soc. dev. Biol.* 40, 275–316 (1982).
4. Raper, J. A., Bastiani, M. J. & Goodman, C. S. *Cold Spring Harb. Symp. quant. Biol.* 48, 587–598 (1983).
5. Goodman, C. S., Bastiani, M. J., Raper, J. A. & Thomas, J. B. in *Molecular Bases of Neural Development* (ed. Cowan, W. M.) (Neuroscience Research Program Press, in the press).
6. Bate, C. M. *J. Embryol. exp. Morph.* 35, 107–123 (1976).
7. Wheeler, W. M. *J. Morph.* 8, 1–148 (1893).
8. Goodman, C. S. & Spitzer, N. C. *Nature* 280, 208–214 (1979).
9. Goodman, C. S., O'Shea, M., McCaman, R. E. & Spitzer, N. C. *Science* 204, 219–222 (1979).
10. Taghert, P. H., Bastiani, M. J., Ho, R. K. & Goodman, C. S. *Devl Biol.* 94, 391–399 (1982).
11. Bastiani, M. J. & Goodman, C. S. *Proc. natn. Acad. Sci. U.S.A.* 81, 1849–1853 (1984).
12. Bastiani, M. J. & Goodman, C. S. in *Cellular and Molecular Approaches to Neuronal Development* (ed. Black, I.) 63–84 (Plenum, New York, 1984).
13. Bastiani, M. J., DuLac, S. & Goodman, C. S. in *Model Neural Networks and Behavior* (ed. Selverston, A.)(Plenum, NewYork, in the press).
14. Raper, J. A., Bastiani, M. J. & Goodman, C. S. *J. Neurosci.* 3, 31–41 (1983).
15. Raper, J. A., Bastiani, M. & Goodman, C. S. *J. Neurosci.* 3, 20–30 (1983).
16. Bastiani, M. J., Raper, J. A. & Goodman, C. S. *J. Neurosci.* (in the press).
17. Raper, J. A., Bastiani, M. J. & Goodman, C. S. *J. Neurosci.* (in the press).
18. Kotrla, K. & Goodman, C. S. *Nature* (in the press).
19. Bentley, D., Keshishian, H., Shankland, M. & Toroian-Raymond, A. *J. Embryol. exp. Morph.* 54, 47–74 (1979).
20. Chang, S., Ho, R. & Goodman, C. S. *Devl Brain Res.* 9, 297–304 (1983).
21. Bownes, M. *J. Embryol exp. Morph.* 33, 789–801 (1975).

DNA sequence and expression of the B95-8 Epstein–Barr virus genome

R. Baer*, A. T. Bankier, M. D. Biggin, P. L. Deininger*, P. J. Farrell†, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Séguin*, P. S. Tuffnell† & B. G. Barrell

MRC Laboratory of Molecular Biology and † Ludwig Institute for Cancer Research, MRC Centre, Hills Road, Cambridge CB2 2QH, UK

The complete (172,282 base pairs) nucleotide sequence of the B95-8 strain of Epstein–Barr virus has been established using the dideoxynucleotide/M13 sequencing procedure. Many RNA polymerase II promoters have been mapped and the mRNAs from these promoters have been assigned to the latent or early/late productive virus cycles. Likely protein-coding regions have been identified and three of these have been shown to encode a ribonucleotide reductase, a DNA polymerase and two surface glycoproteins.

EPSTEIN–BARR virus (EBV) is a human herpesvirus[1] which is endemic in all human populations. Most people are infected with the virus in early childhood and then carry the virus for life. If the initial infection is delayed until adolescence, infectious mononucleosis frequently results. The virus is also linked with certain kinds of cancer. In the malarial belt of Africa, EBV is a contributory factor in the development of Burkitt's lymphoma and in South-East Asia the virus is linked to the high incidence of undifferentiated nasopharyngeal carcinomas (reviewed in refs 2, 3).

The lack of a simple permissive tissue system for EBV makes it difficult to obtain large amounts of virus and hampers genetic analysis. Until recently no genes had been located on the viral genome. Only primate B lymphocytes appear to have receptors for the virus[4–6] but *in vivo* both B lymphocytes and certain (possibly epithelial) cells in the oropharynx become infected[7,8]. There are many different EBV-infected lymphoid cell lines and these derive either from Burkitt's lymphoma explants or from B lymphocytes infected *in vitro* with EBV. When B lymphocytes are infected with EBV they are efficiently immortalized to perpetual growth. The B95-8 cell line was established by infecting marmoset B lymphocytes with EBV from a human with infectious mononucleosis[9]. Only a small proportion of the B95-8 cells support virus production spontaneously; the remainder are considered to be latently infected. The proportion of cells producing EBV can be substantially increased[10] by adding various inducers such as the tumour promoter 12-*O*-tetradecanoylphorbol-13-acetate (TPA).

EBV has a double-stranded DNA genome about 172 kilobases (kb) long which is linear in the virus particle[11,12] but exists as a circular episome inside the nucleus of the infected cell[13,14]. The circularization is mediated by means of multiple direct sequence repeats about 0.5 kb long at the ends of the linear form[15,16] which become joined in the circular form[14]. The genome is further divided into a short and a long unique region by direct sequence repeats (up to 12) of ~3 kb[17–19]. Unlike other herpesviruses, the short (U_S) and long (U_L) unique regions of EBV are maintained in a unique orientation relative to each other. In most EBV-containing cell lines (including B95-8) the majority of EBV DNA is episomal and it is generally thought that most gene expression in such lines is from this form.

The *Eco*RI and *Bam*HI restriction fragments of the virus have been cloned and restriction maps for these enzymes obtained[14,20,21]. By using these cloned fragments as probes in Northern blotting experiments, many viral mRNAs have been approximately localized on the viral genome[22,23]. As a result of hybrid-selected translation experiments on mRNA from EBV-infected cells, many viral proteins have been assigned to regions of the viral genome[24–26]. The genome is thought to be transcribed by the cellular RNA polymerases II and III.

We have now determined the complete DNA sequence of the B95-8 strain of EBV and analysed the sequence for likely coding regions and transcription promoters, splice junctions and poly(A) addition sites. This information is being used to analyse the transcription and gene expression of EBV. Here we show the overall arrangement of possible protein-coding regions and summarize our present knowledge of the transcription and translation of the virus.

DNA sequence analysis

M13 subclone libraries were constructed from suitable *Eco*RI or *Bam*HI fragments of B95-8 EBV by the sonication method[27] using the M13mp8 and M13mp9 vectors[28]. These M13 clones were sequenced randomly by the dideoxynucleotide procedure[29] and the data compiled in a DEC VAX computer using the programs of Staden[30]. About 95% of the sequence of each region was obtained on both strands by the random procedure and the final single-stranded areas or ambiguities were resolved by more directed methods. The methods used have been reviewed by Bankier and Barrell[31]. The sequence of each nucleotide was determined on average 7.3 times and, apart from small parts of certain repetitive regions, determined on both strands. To join up the sequences of adjacent *Eco*RI or *Bam*HI clones, random sequence analysis of a clone overlapping the junction was performed. Restriction maps deduced from the DNA sequence are shown in Fig. 1.

* Present addresses: Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA (R.B.); LSU Medical Center, 1901 Perdido St, New Orleans, Louisiana 70112, USA (P.L.D.); Cellular Regulation Section, Laboratory of Biochemistry, NCI, National Institutes of Health, Bethesda, Maryland 20205, USA (C.S.).

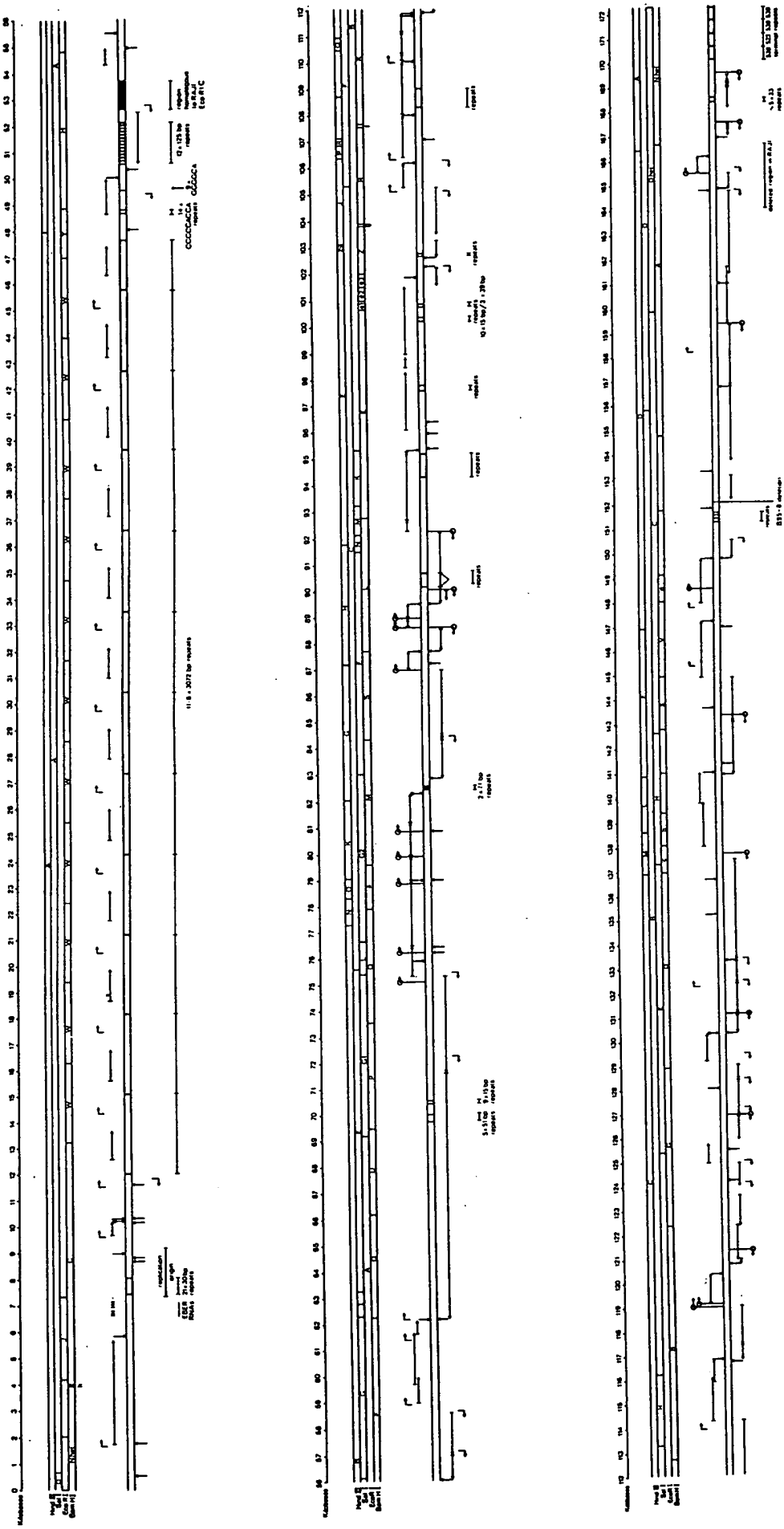

Fig. 1 Arrangement of restriction sites, major open reading frames, promoters and other features in the B95-8 EBV genome. Restriction maps deduced from the DNA sequence are shown for the enzymes *Hin*dIII, *Sal*I, *Eco*RI and *Bam*HI beneath the scale in kb. Position 0 on the scale is the *Eco*RI site between *Eco*RI Dhet and *Eco*RI 1; this is 51 bases to the right of the true left end of the short unique region. Tandemly repeated regions of the genome are shaded. Reading frames are indicated by horizontal arrows and AATAAA (mRNA 3' ends and poly(A) sites) by vertical arrows. ? Marks the rarely used ATTAAA 3' end signal. Promoters confirmed to function in B95-8 cells are shown by the symbol ↱ and promoters surmised from the DNA sequence are indicated by the symbol ↱. The reading frames shown in the diagram were selected on the grounds of length, codon usage, initiator methionine content and position relative to each other and to transcription signals. Confirmed promoters have been mapped by a combination of S_1 mapping and primer extension on RNA from B95-8 cells and in some cases by *in vitro* transcription. Possible promoters have been predicted from position and sequences homologous to the TATA box[53]. The complete DNA sequence and feature table have been deposited with the EMBL Nucleotide Sequence Data Library, Postfach 10.2209, 6900 Heidelberg, FRG.

Table 1 Coordinates of the starts and ends of the reading frames indicated in Fig. 1 and the calculated molecular weights (MW) of the predicted polypeptides

Name	Position		MW	Promoter			Comments
BNRF1	1,736	5,689	142,843				
BCRF1	9,675	10,184	19,914				
BWRF1	12,541	13,689	39,866				12 copies
BYRF1	48,429	49,964	55,173				
BHRF1	54,376	54,948	21,893				
BHLF1	52,557	50,578	66,244	BH-L1	52,786	Early	
BFLF2	56,935	55,982	35,361				
BFLF1	58,525	56,951	57,912				
BFRF1	58,891	59,898	37,632				
BFRF2	59,610	61,580	70,942				
BFRF3	61,456	62,034	19,966				
BPLF1	71,527	62,081	337,973				
BOLF1	75,239	71,523	132,750				
BORF1	75,238	76,329	39,191	BO-R1	75,052	Late	
BORF2	76,407	78,884	93,030	BO-R2	76,198	Early	Ribonucleotide
BaRF1	78,900	79,805	34,358	Ba-R1	78,838	Early	reductase
BMRF1	79,899	81,110	43,373	BM-R1	79,871	Early	
BMRF2	81,118	82,188	39,515	BM-R2	80,811	Late	
BMLF1	84,122	82,746	51,347				
BSLF2	84,288	84,229	2,162				
BSLF1	86,881	84,260	98,040				
BSRF1	86,924	87,577	23,861	BS-R1	86,918	Late	
BLLF2	88,474	87,641	30,952	BL-L3	88,481	Early	
BLRF1	88,547	88,852	10,944	BL-R1	88,539	Late	
BLRF2	88,925	89,410	17,687	BL-R2	88,896	Late	
BLLF1a	92,153	89,433	94,431	BL-L1	92,158	Late	gp350
BLLF1b	92,153	89,433	75,171				gp220
BLLF3	90,013	89,569	16,652	BL-L2	90,020	Early	
BLRF3	92,243	92,599	12,832				
BERF1	92,646	95,162	92,314				Homologous with
BERF2a	95,353	95,721	13,186				BERF2b, BERF4
BERF2b	95,725	98,244	92,769				Homologous with
BERF3	98,323	98,766	16,717				BERF1, BERF4
BERF4	98,805	101,420	95,543				Homologous with
BZLF2	102,116	101,448	25,257				BERF1, BERF2b
BZLF1	103,155	102,556	21,482				
BRLF1	105,183	103,369	66,594				
BRLF2	104,989	104,927	2,343				
BRRF1	105,182	106,111	35,319				
BRRF2	106,302	107,912	56,954				
BKRF1	107,950	109,872	56,427				EBNA
BKRF2	109,958	110,368	15,080				
BKRF3	110,275	111,117	31,606				
BKRF4	111,107	111,784	24,837				
BBLF4	114,259	111,833	89,853				
BBRF1	114,204	116,042	68,456				
BBRF2	115,843	116,781	34,916				
BBLF3	117,386	116,784	22,605				
BBLF2	119,080	117,416	60,364				
				BB-R1	119,014	Late	
BBRF3	119,137	120,351	45,792	BB-R3	119,133	Late	
BBLF1	120,974	120,750	8,470	BB-L1	121,300	Late	
BGLF5	122,341	120,932	52,666				
BGLF4	123,692	122,328	51,291				
BGLF3	124,939	123,944	37,708				
BGRF1	124,938	125,912	36,462				
BGLF2	126,873	125,866	36,888	EE-L8	126,895	Late	
BGLF1	128,374	126,854	54,462				
BDLF4	129,021	128,347	25,448				
BDRF1	129,188	130,348	42,626				
BDLF3	131,066	130,365	23,791	EE-L4	131,073	Late	Glycoprotein?
BDLF2	132,389	131,130	46,168				
BDLF1	133,307	132,403	33,624				
BcLF1	137,466	133,324	153,916	EH-L1	137,676	Late	
BcRF1	137,862	139,715	68,711				
BTRF1	139,642	140,916	46,711				
BXLF2	143,036	140,919	78,321	EC-L2	143,274	Late	
BXLF1	144,861	143,041	67,193				
BXRF1	144,860	145,603	27,063				
BVRF1	145,416	147,125	62,461				
BVRF2	147,927	149,741	64,102				
BdRF1	148,707	149,741	36,127	EC-R1	148,651	Late	
BILF2	150,525	149,782	27,076				Glycoprotein?
BILF1	153,099	152,164	34,519				Membrane protein?
BALF5	156,746	153,702	113,419				DNA polymerase
BALF4	159,322	156,752	95,640	EC-L1	159,337	Late	Glycoprotein?
BALF3	161,678	159,312	85,536				
BALF2	164,770	161,387	123,122				
BALF1	165,517	164,858	25,149				
BARF1	165,504	166,166	24,471	ED-R1	165,498	Early	
BNLF2b	167,303	167,001	11,449				
BNLF2a	167,486	167,307	6,540	ED-L2	167,495	Early	
BNLF1c	168,966	168,163	28,851				
BNLF1b	169,129	169,043	3,212				
BNLF1a	169,474	169,208	9,942	ED-L1	169,517	Latent	

BLLF1a and b are overlapping co-terminal reading frames, BLLF1b having a central portion spliced out. The locations of the approximate transcription start points of the confirmed promoters and the stage of expression of mRNAs from these promoters are given. Reading frames identified as coding for known proteins are indicated. Reading frames rich in the glycosylation site sequence (N-X-T/S) are also noted, together with a possible membrane protein, which has many hydrophobic amino acid residues.

The complete DNA sequence, together with a table showing the positions of the features shown in Fig. 1, has been deposited with the EMBL database; for reasons of space the sequence and feature table cannot be shown here. We have previously published the sequences of the *Eco*RI Dhet, *Eco*RI C, *Bam*HI B, *Bam*HI L, *Bam*HI a and part of the *Bam*HI M regions together with some analysis of their transcription[32–38].

Most of the sequence we have established is from single *Eco*RI or *Bam*HI clones. The possibility that cloning artefacts may have arisen in the construction and maintenance of the *Eco*RI and *Bam*HI libraries cannot be excluded, but there is no reason to believe that such artefacts have occurred. The overlap between the *Eco*RI Dhet and *Eco*RI I fragments was obtained by sequencing the corresponding region of EBV strain, M-ABA, which has a single base change in the *Eco*RI recognition sequence.

There are some well documented areas of the EBV genome in which some strains have deletions relative to most other strains: one of these is in the *Bam*HI WYH region and another is in the *Eco*RI C region. B95-8 does not have a deletion in the *Bam*HI WYH region; the P3HR1 and Daudi strains, however, are missing bases 45,644 to 52,450 and 45,415 to 52,824 of our sequence respectively[39,40]. There is a deletion of ~13.6 kb in B95-8 relative to most strains and we have determined that it lies between bases 152,012–152,013 by comparison with the sequence of the corresponding fragment of Raji DNA[33]. Similarly, a deletion of 2,658 base pairs (bp) in the *Eco*RI Dhet fragment of Raji DNA removes bases 163,978 to 166,635 of the B95-8 sequence.

Repeat sequences

The genome is divided into the two unique regions (U_S and U_L) by the major internal repeat. We have avoided an earlier nomenclature in which the genome is subdivided into smaller unique regions by other repetitive sequences because there are many more repeat sequences in the virus than was previously thought and these do not appear to separate functional domains of the virus.

In the *Eco*RI Dhet clone that we sequenced (derived from the circular form of the virus) there were four copies of the terminal repeat, three of 538 bp and one of 523 bp[32]. In the 3.07-kb major internal repeat the sequence of only one of the *Bam*HI W clones was determined and for the moment we have assumed that the repeats are identical. The number of copies of this repeat (11.6 copies in the B95-8 strain) has been taken from previous work[41]. Our sequence of *Bam*HI W is identical to that determined by Jones and Griffin[42] but differs in two places from the sequence of Cheung and Kieff[43].

Repetitive regions are scattered throughout the genome. Some repeats are found in likely coding regions and it is known in some cases that the repeats are actually encoded into protein: one of the most striking of these is the *Bam*HI K Epstein–Barr nuclear antigen (EBNA) Gly-Ala repeat[44] in the BKRF1 reading frame (BKRF1 is explained below). In this repeat and several others there is no degeneration (third position variation) in the repeat sequences which are coding. Some mechanism or constraint apart from simply coding for protein may prevent these repeat sequences from drifting.

Interpretation of the sequence

The DNA sequence has been analysed[45] for transcription promoters, major open reading frames and possible polyadenylation/mRNA processing sites. The reading frames, promoters and AATAAA signals are shown in Fig. 1 together with other features of the sequence such as repetitive sequences. In our nomenclature, reading frames and promoters are preceded by the abbreviated name of the restriction fragment in which they start translation or transcription. A promoter starting transcription in *Bam*HI K is preceded by BK-, one starting in *Eco*RI C is preceded by EC-. This is followed by L or R depending on whether the promoter or reading frame is leftward or rightward on the standard map. Promoters are then simply numbered;

reading frames have F and a number. Thus, BN-R1 would be a rightward promoter starting transcription in *Bam*HI N and BCLF2 would be a leftward reading frame beginning in *Bam*HI C. In many cases (see below) we have demonstrated promoters to be functional in B95-8 cells but some promoters are still only predicted from the DNA sequence.

On grounds of size alone it is likely that all or large parts of the reading frames shown are expressed as protein. Their positions with respect to each other and to promoter and polyadenylation/RNA processing sites strengthen this argument. Because of the high G+C content (59.94% in B95-8) we have been able to use codon usage analysis to further justify many of the reading frames[33]. There are 84 unique major open reading frames shown in Fig. 1 and the coordinates and sizes of the reading frames and locations of known transcription starts and various other features are shown in Table 1. The sequence is numbered from the *Eco*RI site separating *Eco*RI Dhet from *Eco*RI I.

Generally the arrangement of coding regions in the unique regions is economical, particularly in the large unique region. Often there are few or no bases between reading frames and some promoters are found to lie in the coding regions of the adjacent genes. This economical arrangement in the unique regions contrasts with the large repetitive regions of the virus. The number of major internal repeat units is variable in different strains[17,18,46,47] and tends to compensate for the deletions that certain strains have; this may indicate a packing constraint on the size of the virus.

Gene expression

The total number of proteins expressed from the EBV genome is unknown. Of the 84 major open reading frames, some may be spliced together, tending to reduce the total number of proteins expressed by the virus, while alternate splicing patterns would increase the number of proteins. Some proteins (discussed below) are thought to correspond to the major antigens of the virus, EBNA, MA (membrane antigen) and the capsid proteins (VCA). Other classically identified antigens such as the EA (early antigen) group and LYDMA (lymphocyte-detected membrane antigen), remain to be identified as polypeptide chains. A large number of EBV-specific polypeptides have been mapped to particular regions of the virus by hybrid-selected translation[24–26], though in some regions of the virus no proteins have yet been detected by this method. For example, the two very large leftward reading frames BPLF1 and BOLF1 would encode polypeptides of molecular weight 338,000 and 133,000, respectively. However, no proteins have been reported as originating from this substantial region of the virus, though the fact that the reading frames are so large makes it fairly evident that they are expressed.

The EBV genome is thought to be transcribed by the cellular RNA polymerases II and III. RNA polymerase III transcribes the two Epstein–Barr early region (EBER) RNA genes in the small unique region[48,49]. There is no evidence of any tRNA genes in EBV and our computer searches of the sequence have revealed none. The function of the EBER RNAs is not directly known but they will substitute for the adenovirus VA RNAs in an adenovirus infection[50]. The EBER RNA genes are localized next to the maintenance origin of replication of EBV mapped by Yates *et al.*[51].

We have searched for RNA polymerase II (pol II) promoters by a combination of identifying 5′ ends of EBV mRNAs using S_1 mapping and primer extension experiments and by *in vitro* transcription of EBV DNA fragments in a HeLa whole-cell extract[52]. So far, 24 pol II promoters have been detected but the final total is expected to be more than twice this number. The location of the transcription starts of these promoters (which are all confirmed to function in B95-8 cells) are given in Table 1. All the promoters detected so far have sequences about 30 bases upstream of their transcription starts that are homologous with the TATAAAA sequence[53].

RNAs have been classified as expressed in the latent cycle or early or late productive cycles[22,23]. We have compared the levels of RNAs in B95-8 cells either not treated or treated with TPA. Although the control B95-8 cells produce virus, the levels of productive cycle RNAs are low in control cultures and increase very dramatically on TPA treatment so, in practice, it is easy to distinguish latent cycle RNAs from productive cycle ones in this system. Phosphonoacetic acid (PAA) is used as an inhibitor of viral DNA synthesis[54] to distinguish between early productive cycle RNAs and late productive cycle ones. PAA prevents the TPA induction of late RNAs but not of early RNAs.

It is unknown whether the regulated expression of EBV genes is controlled at the transcriptional level or through RNA processing or stability; there is a precedent for transcriptional regulation in herpes simplex virus (HSV)[55]. We previously noted homologous DNA sequences upstream of some EBV promoters[37,38] which might be used to regulate those genes at the transcriptional level. All the promoters which had those sequences give rise to late RNAs but only a few of the late promoters have the sequences, so it remains to be seen whether they are functionally important. The promoters are classified as latent, early or late in Table 1 though at present this classification refers to when the corresponding RNA is observed, rather than proven regulation of the promoter activity.

Genes active in latently infected cells

Significant levels of mRNA are expressed from three regions of the genome in latently infected cells in addition to the pol III EBER RNA transcripts[23,56,57]: one of these regions is the *Bam*HI WYH region where mRNA is latently transcribed rightward possibly from the *in vitro* promoter[58] in *Bam*HI W (transcription start at position 45,104) or from the small unique region. A mature latent mRNA of 3.0 kb containing exons in *Bam* W, Y and H has been described[58]. We have confirmed that there is a strong *in vitro* promoter starting transcription at position 45,104 but have been unable to show that it operates in B95-8 cells. Our preliminary Northern blotting experiments in the *Bam* W, Y, H region reveal two spliced rightward latent RNAs of 2.4 and 3.7 kb. The deletion in P3HR1 which may account for the nontransforming phenotype of this strain affects the region where these mRNAs map as well as the early RNA transcribed across the 125-bp repeats, encoding BHLF1[39].

A second latently transcribed region is *Bam*HI K. A 3.7-kb latent RNA hybridizes with *Bam*HI K which contains a simple repeat sequence 708 bases long, consisting only of the triplets GGA, GGG and GCA[59]. This repeat sequence lies in the BKRF1 reading frame and codes for an amino acid sequence consisting of only Gly and Ala. This region has been conclusively shown to code for the nuclear antigen EBNA-1 (refs 44, 60). The molecular weight of the BKRF1 reading frame is 56,427 which is lower than the observed 68,000–85,000 of the EBNA-1 protein. The molecular weight of EBNA varies in a strain-specific manner according to the number of repeats in the BKRF1 frame[61]. The location of the promoter for latent transcription of the EBNA-1 gene is not yet known but may lie several kilobases upstream of the start of the BKRF1 frame[59]. The observation[60] that a protein very similar to EBNA is produced in response to transfection of a *Bam*HI–*Hin*dIII fragment (positions 107,565–110,491) implies that BKRF1 accounts for most of the coding sequence. Another EBNA (EBNA-2) has been identified, though the gene for this has not yet been accurately localized on the genome map[44]. A third nuclear antigen appears to map to the *Bam*HI M region[62].

The most abundantly transcribed EBV mRNA in latently infected cells is a 2.8-kb RNA encoded by *Eco*RI Dhet[23,56,57] which has been correlated with a latent active leftward promoter[38] at position 169,513 (ED-L1). In the latent virus cycle the RNA from this promoter is spliced and most of the mRNA is composed of an exon containing the BNLF1c reading frame and terminating at the poly(A) addition site at 166,950. This mRNA probably codes for a 42,000-molecular weight membrane protein[32]. The 5′ end of this mRNA may be different in the productively infected cell.

Early productive cycle genes

Productive cycle RNAs are induced in B95-8 cells by treatment with TPA and induction of early RNAs is not inhibited by blocking viral DNA synthesis with PAA. The functions of three reading frames in B95-8 EBV have been identified by comparing the protein-coding sequences of all the reading frames in EBV with a library of known protein sequences. RNAs encoding these reading frames are expressed in the early productive cycle. We found that the reading frame BALF5 is similar to the HSV DNA polymerase (J. Quinn and D. McGeoch, personal communication) and the reading frames BORF2 and BaRF1 are similar (112 amino acids out of 301) to the HSV ribonucleotide reductase gene region[36]. The EBV reading frame BXLF1 has a small region of homology (19 amino acids out of 35) with the HSV2 thymidine kinase gene[63]. This sequence, however, may represent a nucleotide-binding site[64] rather than necessarily imply a thymidine kinase function for BXLF1.

Late productive cycle genes

The major component of the virus capsid is a 160,000 (160K)-molecular weight protein and a similar-sized protein has been observed by hybrid-selected translation using the *Eco*RI E fragment[22]. This protein may be encoded by the BcLF1 reading frame which would give a 154K protein.

The membrane antigen (MA) of EBV contains several proteins including gp350/300, gp250/200, p140 and gp85: three of these are glycoproteins (gp). Antibodies to MA and to the gp350/300 neutralize viral infectivity in tissue culture[65,66]. The sequences of these proteins are of interest, therefore, in the development of synthetic vaccines against EBV infection. Hummel *et al.*[67] have mapped gp350/300 and gp250/200 to the *Bam*HI L region by hybrid-select translation and suggested that they may be expressed from overlapping reading frames because they have peptides in common. By Northern blotting and S_1 mapping[35], it has been shown that there are two co-terminal late RNAs containing the BLLF1 reading frame, the smaller having most or all of an internal repetitive region removed by splicing[35]. It is proposed that gp350/300 is expressed from a 2.8-kb mRNA and gp250/200 is expressed from a 2.2-kb mRNA which is spliced, removing the repetitive region[35].

Although the location of the gene for the gp85 protein is unknown, two obvious candidates are the BALF4 and the BDLF3 reading frames as they are about the right size and contain many potential glycosylation sites. The non-glycosylated membrane protein p140 is 143K in size and a similar-sized viral protein was mapped to the short unique region[22]; this protein is probably encoded by BNRF1.

Conclusions

Our structural approach to the biology of EBV, making use of the complete DNA sequence, has been particularly useful because of the lack of viral genetics and because of technical obstacles to working with the virus in tissue culture. The library of over 6,000 characterized M13 clones covering the genome obtained from the sequencing programme is proving most useful in analysing the gene expression of the virus. By using the M13 clones as probes for S_1 mapping and Northern blotting experiments, we are now constructing a detailed transcription map of the virus.

One of the most interesting features of EBV is its ability to immortalize B lymphocytes. Only a few regions of the genome are expressed in the latently infected lymphocyte: some of these may be involved in maintenance of the viral DNA and one at least is presumably involved in the immortalization. Identifying the immortalizing function may be useful both technically for making human monoclonal antibody lines that do not secrete EBV and with respect to the involvement of the virus in oncogenesis. EBV does not seem to be the proximal cause of Burkitt's lymphoma (reviewed in ref. 68) but is nevertheless a contributory risk factor. The fact that virtually every South-East Asian undifferentiated nasopharyngeal carcinoma carries EBV DNA implies a link between the virus and this disease. A region of the EBV genome which immortalizes epithelial cells has been identified[69] but immortalization of lymphocytes may require other EBV genes. At present it seems that an understanding of the role of the virus in these diseases will probably derive from a future investigation of the detailed molecular biology of EBV.

We thank B. E. Griffin and J. Arrand for the B95-8 and Raji *Eco*RI and *Bam*HI cloned restriction fragment libraries, and G. Bornkamm for the terminal M-ABA *Eco*RI clone. We also thank G. Bornkamm, B. E. Griffin, G. S. Hayward, S. D. Hayward, E. Kieff, G. Klein, D. J. McGeogh, G. Miller, M. Perricaudet, J. Pagano, J. Strominger, B. Sugden, H. Wolf and J. Yates who communicated their work to us before publication. During this study G.S.H. was supported by an Usher fellowship, P.L.D. by NATO, G.H. by the MRC, R.B. by the Damon Runyon–Walter Winchell Cancer Fund and C.S. by the MRC of Canada.

Received 28 February; accepted 1 May 1984.

1. Epstein, M. A., Achong, B. G. & Barr, Y. M. *Lancet* **i**, 702–703 (1964).
2. Roizman, B. (ed.) *The Herpesviruses, The Viruses* Vol. 1 (Plenum, New York, 1982).
3. Epstein, M. A. & Achong, B. G. (eds) *The Epstein–Barr Virus* (Springer, Berlin, 1979).
4. Henle, W., Diehl, V., Kohn, G., zur Hausen, H. & Henle, G. *Science* **157**, 1064–1065 (1967).
5. Pope, J. H., Horne, M. K. & Scott, W. *Int. J. Cancer* **3**, 857–866 (1968).
6. Jondal, M. & Klein, G. *J. exp. Med.* **138**, 1365–1378 (1973).
7. Morgan, D. G., Niederman, J. C., Miller, G., Smith, H. W. & Dowaliby, J. M. *Lancet* **ii**, 1154–1157 (1979).
8. Lemon, S. M., Hutt, L. M., Shaw, J. E., Li, J. L. H. & Pagano, J. S. *Nature* **268**, 268–270 (1977).
9. Miller, G., Shope, T., Lisco, H., Stitt, D. & Lipman, M. *Proc. natn. Acad. Sci. U.S.A.* **69**, 383–387 (1972).
10. zur Hausen, H., Hecker, E., O'Neill, F. J. & Freese, U. K. *Nature* **272**, 373–375 (1978).
11. Pritchett, R. F., Hayward, S. D. & Kieff, E. *J. Virol.* **15**, 556–569 (1975).
12. Hayward, S. D. & Kieff, E. *J. Virol.* **23**, 421–429 (1977).
13. Lindahl, T. *et al. J. molec. Biol.* **102**, 511–530 (1976).
14. Dambaugh, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 2999–3003 (1980).
15. Given, D., Yee, D., Griem, K. & Kieff, E. *J. Virol.* **30**, 852–862 (1979).
16. Kintner, C. R. & Sugden, B. *Cell* **17**, 661–671 (1979).
17. Given, D. & Kieff, E. *J. Virol.* **31**, 315–324 (1979).
18. Rymo, L. & Forsblom, S. *Nucleic Acids Res.* **5**, 1387–1402 (1978).
19. Hayward, S. D., Nogee, L. & Hayward, G. S. *J. Virol.* **33**, 507–521 (1980).
20. Skare, J. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3860–3864, 7510 (1980).
21. Arrand, J. R. *et al. Nucleic Acids Res.* **9**, 2999–3014 (1981).
22. Hummel, M. & Kieff, E. *J. Virol.* **43**, 262–272 (1982).
23. Weigel, R. & Miller, G. *Virology* **125**, 287–298 (1983).
24. Hummel, M. & Kieff, E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5698–5702 (1982).
25. Seibl, R. & Wolf, H. *Virology* (submitted).
26. Cohen, L. K., Spech, S. H., Roberts, B. E. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
27. Deininger, P. *Analyt. Biochem.* **129**, 216–223 (1983).
28. Messing, J. & Vieira, J. *Gene* **19**, 269–276 (1982).
29. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
30. Staden, R. *Nucleic Acids Res.* **10**, 4731–4751 (1982); **12**, 499–503 (1984).
31. Bankier, A. T. & Barrell, B. G. *Techniques in the Life Sciences* Vol. B5 (Elsevier, Ireland, 1983).
32. Bankier, A. T. *et al. Molec. Biol. Med.* **1**, 425–445 (1983).
33. Bankier, A. T., Deininger, P. L., Farrell, P. J. & Barrell, B. G. *Molec. Biol. Med.* **1**, 21–45 (1983).
34. Seguin, C. S., Farrell, P. J. & Barrell, B. G. *Molec. Biol. Med.* **1**, 369–392 (1983).
35. Biggin, M., Farrell, P. J. & Barrell, B. G. *EMBO J.* **3**, 1083–1090 (1984).
36. Gibson, T., Stockwell, P., Ginsberg, M. & Barrell, B. G. *Nucleic Acids Res.* (in the press).
37. Farrell, P. J., Deininger, P. L., Bankier, A. T. & Barrell, B. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1565–1569 (1983).
38. Farrell, P. J., Bankier, A., Seguin, C., Deininger, P. & Barrell, B. G. *EMBO J.* **2**, 1331–1338 (1983).
39. Jeang, K. & Hayward, S. D. *J. Virol.* **48**, 135–148 (1983).
40. Jones, M., Foster, L., Sheedy, T. & Griffin, B. E. *EMBO J.* **3**, 813–821 (1984).
41. Hayward, S. D., Lazarowitz, S. G. & Hayward, G. S. *J. Virol.* **43**, 201–212 (1982).
42. Jones, M. D. & Griffin, B. E. *Nucleic Acids Res.* **11**, 3919–3937 (1983).
43. Cheung, A. & Kieff, E. *J. Virol.* **44**, 286–294 (1982).
44. Hennessy, K. & Kieff, E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5665–5669 (1983).
45. Staden, R., *Nucleic Acids Res.* **12**, 521–538 (1984).
46. Raab-Traub, N., Pritchett, R. & Kieff, E. *J. Virol.* **27**, 388–398 (1978).
47. Delius, H. & Bornkamm, G. W. *J. Virol.* **27**, 81–89 (1978).
48. Arrand, J. R. & Rymo, L. *J. Virol.* **41**, 376–379 (1982).
49. Rosa, M. D., Gottlieb, E., Lerner, M. R. & Steitz, J. A. *Molec. cell. Biol.* **1**, 785–796 (1981).
50. Bhat, R. A. & Thimmappaya, B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4789–4793 (1983).
51. Yates, J., Warren, N., Reisman, D. & Sugden, B. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
52. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3855–3859 (1980).
53. Corden, J. *et al. Science* **209**, 1406–1414 (1980).
54. Summers, W. C. & Klein, G. *J. Virol.* **18**, 151–155 (1976).
55. Herz, C. & Roizman, B. *Cell* **33**, 145–151 (1983).
56. Kieff, E. *et al. J. infect. Dis.* **146**, 506–517 (1982).
57. Rymo, L. *J. Virol.* **32**, 8–18 (1979).
58. Van Santen, V., Cheung, A., Hummel, M. & Kieff, E. *J. Virol.* **46**, 424–433 (1983).
59. Heller, M., van Santen, V. & Kieff, E. *J. Virol.* **44**, 311–320 (1982).
60. Fischer, D. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 43–47 (1984).
61. Hennessy, K., Heller, M., van Santen, V. & Kieff, E. *Science* **220**, 1396–1398 (1983).
62. Grogan, E. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 7650–7653 (1983).
63. Swain, M. A. & Galloway, D. A. *J. Virol.* **46**, 1045–1050 (1983).
64. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. *EMBO J.* **1**, 945–951 (1982).
65. Hofmann, G. J., Lazarowitz, S. G. & Hayward, S. D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2979–2983 (1980).
66. Thorley-Lawson, D. A. & Poodry, C. A. *J. Virol.* **43**, 730–736 (1982).
67. Hummel, M., Thorley-Lawson, D. A. & Kieff, E. *J. Virol.* **49**, 413–417 (1984).
68. Klein, G. *Cell* **32**, 311–315 (1983).
69. Griffin, B. E. & Karran, L. *Nature* **309**, 78–82 (1984).

JOURNAL OF VIROLOGY, Aug. 1992, p. 5030–5039
0022-538X/92/085030-10$02.00/0

Vol. 66, No. 8

trans-Acting Requirements for Replication of Epstein-Barr Virus ori-Lyt

ELIZABETH D. FIXMAN,[1,2] GARY S. HAYWARD,[1] AND S. DIANE HAYWARD[1,2]*

The Virology Laboratories, Department of Pharmacology and Molecular Sciences,[1] and Department of Neurology,[2] Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205*

Received 21 February 1992/Accepted 1 May 1992

Epstein-Barr virus (EBV) utilizes a completely different mode of DNA replication during the lytic cycle than that employed during latency. The latency origin of replication, ori-P, which functions in the replication of the latent episomal form of the EBV genome, requires only a single virally encoded protein, EBNA-1, for its activity. During the lytic cycle, a separate origin, ori-Lyt, is utilized. Relatively little is known about the *trans*-acting proteins involved in ori-Lyt replication. We established a cotransfection-replication assay to identify EBV genes whose products are required for replication of ori-Lyt. In this assay, a *Bam*HI-H plasmid containing ori-Lyt was replicated in Vero cells cotransfected with the *Bam*HI-H target, the three EBV lytic-cycle transactivators Zta, Rta, and Mta, and the EBV genome provided in the form of a set of six overlapping cosmid clones. By removing individual cosmids from the cotransfection mixture, we found that only three of the six cosmids were necessary for ori-Lyt replication. Subcloning of the essential cosmids led to the identification of six EBV genes that encode replication proteins. These genes and their functions (either known or predicted on the basis of sequence comparison with herpes simplex virus) are BALF5, the DNA polymerase; BALF2, the single-stranded DNA-binding protein homolog; BMRF1, the DNA polymerase processivity factor; BSLF1 and BBLF4, the primase and helicase homologs; and BBLF2/3, a potential homolog of the third component of the helicase-primase complex. In addition, ori-Lyt replication in this cotransfection assay was also dependent on one or more genes provided by the EBV *Sal*I-F fragment and on the three lytic-cycle transactivators Zta, Rta, and Mta.

Epstein-Barr virus (EBV), like all herpesviruses, has both a latent state and a lytic replicative cycle. In latently infected B cells, multiple copies of the viral genome are maintained predominantly as nucleosome-covered episomes that are replicated in synchrony with cell division (73). Latency replication proceeds from ori-P, which is composed of two domains (72). Region I, the family of repeats, contains 20 tandemly repeated binding sites for EBNA-1 (57) and functions as an EBNA-1-dependent enhancer whose activity is important for both ori-P replication and transcriptional activation of the *Bam*HI-C latency promoter (58, 63, 71). Region II, the dyad symmetry, contains two pairs of overlapping EBNA-1-binding sites (57) and is the site of initiation of latency replication (26). EBNA-1 is the only virally encoded protein required for replication of the episomal EBV genome, all other proteins, including the DNA polymerase, being provided by the host cell (74).

Lytic EBV replication occurs in mucosal epithelial cells of the oropharynx and genital tract (60) and can be activated in latently infected B cells in culture by treatment with phorbol esters (77), by superinfection with the P3HR-1 strain which carries defective rearranged viral genomes (53), or by introduction of the EBV Zta transactivator (13). Because of the limitations imposed by the lack of an EBV-infectable epithelial culture system, information on EBV lytic viral replication has been obtained predominantly in B-cell cultures undergoing reactivation. In this system, the transition from latency to a lytic replicative cycle is mediated by three viral regulatory proteins, the Zta (BZLF1, EB1, or ZEBRA) and Rta (BRLF1) transcriptional transactivators (9, 12, 13, 31, 34, 48) and Mta (BMLF1), which has a posttranscriptional mechanism of action (5, 39, 49). The concerted action of these three proteins results in activation of the complete cascade of early and late EBV gene expression.

Lytic DNA replication proceeds from a separate origin, ori-Lyt, and results in 100- to 1,000-fold amplification of the genome via concatemeric intermediates (18, 32, 59). In the prototype EBV genome, there are two copies of ori-Lyt, one in DS-L and one in DS-R. However, one copy is sufficient for lytic-cycle replication as exemplified by the EBV strain B95-8, which contains only DS-L. ori-Lyt covers 690 bp and can be divided into three essential domains. (i) The first domain is the promoter and leader of the BHLF1 gene whose transcript is the most abundantly synthesized of the lytic-cycle mRNAs (38). (In the DS-R origin, this promoter controls the related *Pst*I repeat gene.) The BHLF1 promoter contains four binding sites for the Zta transactivator and is strongly Zta responsive in transient expression assays (47, 48). (ii) The second domain is a central 225-bp region whose prominent features include two related AT-rich palindromes of 18 and 20 bp and an adjacent polypurine-polypyrimidine tract. Elements of this type destabilize helical structure and may serve as sites for the initiation or transmission of localized unwinding in origins of replication (42, 68). (iii) The third domain is a powerful enhancer element that responds to the Rta transactivator and contains two binding sites for Rta and one for Zta (14, 31, 48). Hammerschmidt and Sugden (32) found that origin function was retained when this enhancer was replaced with the enhancer from the human cytomegalovirus (HCMV) major immediate-early gene, suggesting that it was enhancer function per se that was provided by this region rather than a contribution involving specific protein interactions.

* Corresponding author.

The lytic origins of replication have been identified for a number of herpesviruses, including HCMV and simian cytomegalovirus (SCMV) (2, 33), varicella-zoster virus (VZV) (17), pseudorabies virus (43), equine herpesvirus 1 (4), Marek's disease virus (7), and herpes simplex virus (HSV) (51, 62, 67). Of these herpesviruses, replication of HSV has been the most extensively characterized. The HSV viral proteins involved in DNA replication were originally recognized through genetic studies (reviewed in reference 66). Subsequently, a complete set of seven essential genes was identified by Challberg et al. (8, 52, 70) using transient cotransfection replication assays.

We established a transient replication assay in Vero cells to determine exactly which EBV proteins are required to replicate an ori-Lyt-containing target. Utilizing this system, we identified six essential EBV replication genes. In addition to these genes, ori-Lyt replication in the transient cotransfection assay was also dependent on one or more genes provided by EBV *Sal*I-F and on the Zta, Rta, and Mta lytic-cycle transactivators.

MATERIALS AND METHODS

Cells and DNA transfections. Vero cells were maintained in Dulbecco modified Eagle medium plus 10% fetal calf serum. One day before transfection, 10^6 cells were plated in 100-mm dishes. Four hours before transfection, the medium was replaced with 10 ml of Dulbecco modified Eagle medium containing 10% fetal calf serum and antibiotics. DNA was transfected by using the calcium phosphate procedure originally described by Graham and van der Eb (30) as modified by Chen and Okayama (11). DNA (12.5 to 14.5 μg) was diluted with water to a total volume of 450 μl. To this was added 50 μl of 2.5 M $CaCl_2$ and 500 μl of 2× BES [*N,N'*-bis(2-hydroxyethyl)-2-aminoethanesulfonic acid]-buffered saline (50 mM BES [pH 6.95], 280 mM NaCl, 1.5 mM Na_2HPO_4). This cocktail was incubated at room temperature for 20 min and then added dropwise to the cells. After incubation for 20 h at 35°C in 3.5% CO_2, the medium was removed, the cells were washed once with phosphate-buffered saline (PBS), and fresh medium containing antibiotics was added. The cells were harvested after a further 72-h incubation.

DNA replication assay. Cell pellets were resuspended in 100 μl of PBS and then lysed in 2 ml of buffer containing 10 mM Tris-Cl (pH 8.0), 10 mM EDTA, 2% sodium dodecyl sulfate (SDS), and 100 μg of proteinase K per ml (8). After overnight incubation at 37°C, the samples were diluted to 4 ml with water. Sodium acetate (pH 5.2) was added to a final concentration of 0.3 M, and the samples were extracted with phenol-chloroform and chloroform and then subjected to ethanol precipitation. The DNA pellets were resuspended in 450 μl of water, treated with RNase, ethanol precipitated a second time, and finally resuspended in 200 μl of water. DNA (5.0 μg) was cut with 10 U of *Bam*HI or 10 U of *Bam*HI and 12 U of *Dpn*I (Boehringer Mannheim) in a reaction volume of 100 μl overnight at 37°C. To check the *Dpn*I activity, we incubated 4 μl of the *Dpn*I reaction digest with 500 ng of pUC19 DNA overnight at 37°C. Complete cutting of the pUC19 DNA was taken to indicate that the experimental DNA was also completely digested. The cellular DNA was then separated by electrophoresis on a 1% gel and transferred to a NYTRAN membrane following the method of Southern (61) as modified by the manufacturer (Schleicher & Schuell, Inc., Keene, N.H.). After the transfer, the NYTRAN membrane was neutralized in 2× SSC (0.3 M NaCl, 0.03 M sodium citrate) and baked under vacuum at 80°C for 1 h. The NYTRAN membrane was prehybridized for 2 h in buffer containing 1% SDS, 5 mg of nonfat dried milk per ml, 0.5 mg of heparin per ml, 0.2 mg of sonicated, denatured salmon sperm DNA per ml, 60 mg of polyethylene glycol 8000, 5× SSPE (750 mM NaCl, 50 mM NaH_2PO_4, 5 mM Na_2EDTA), and 10% formamide (75). The membrane was then incubated overnight at 60°C with 2×10^6 to 5×10^6 cpm of EBV B95-8 *Bam*HI-H, M-ABA *Bam*HI-H (pPDL7), or pBR322 DNA probe per ml and radiolabeled by random priming (25) to a specific activity of 5×10^8 cpm/μg, after which the filters were washed twice in 0.1× SSC–0.1% SDS at 65°C for 45 min and exposed to X-ray film for 24 h at −80°C with an intensifying screen.

Plasmid constructions. The target *Bam*HI-H plasmid (pSL77) used in the transient replication assays has been described previously (49), as have the effector DNA plasmids, pPL17 (Zta), pMH48 (Rta), and pTS6 (Mta) (12, 34, 47). The variant ori-Lyt target, pPL2A, contains the *Bgl*II-C fragment from M-ABA (pM-B2-C [56]) in which the BHLF1 open reading frame has been disrupted by deletion of the *Not*I repeats. To express the B95-8 BMRF1 gene, a 3,026-bp *Bcl*I-*Eco*RI fragment from *Bam*HI-M was ligated into a *Bgl*II-*Eco*RI-cleaved SV2neo derivative, pGH52, which contains a *Hin*dIII-*Bgl*II-*Hin*dIII linker inserted at the *Hin*dIII site of SV2neo. A BSLF1 expression vector, pDH131, was generated by first ligating a 2,819-bp *Bgl*II-*Bam*HI fragment containing BSLF1 into the *Bam*HI site of pUC18 and then transferring the open reading frame as a *Hin*dIII-*Bam*HI fragment into *Hin*dIII-*Bam*HI-cleaved pSV2neo. To express the EBV M-ABA strain BBLF4 gene, 10-mer *Hin*dIII linkers were ligated onto a 3,003-bp *Stu*I fragment containing the entire BBLF4 open reading frame from the cosmid cM301-99 (3, 56), and this fragment was then cloned into the *Hin*dIII site of the pBR322 derivative, pGH59. This plasmid was cut with *Rsr*II, the overhang was filled in with the Klenow DNA polymerase, and the DNA was recut with *Eco*RI. The resulting fragment containing BBLF4 was then ligated into the *Sma*I-*Eco*RI-cleaved SCMV Colburn IE94 expression plasmid, pGH70, to generate pEF54A. To express the M-ABA strain BBLF2 and BBLF3 open reading frames, we inserted the 3,052-bp *Asp*718-*Sal*I fragment from the cosmid cM301-99 into *Asp*718-*Sal*I-cleaved pUC19, generating pEF58. This plasmid was cut to completion with *Sal*I and partially digested with *Bgl*II, and the resulting fragment was cloned into the *Bam*HI-*Sal*I-cleaved SCMV Colburn IE94 expression plasmid pGH177, generating EF59A.

To express the M-ABA strain BALF2 open reading frame, we moved a 3,912-bp *Bgl*II-*Eco*RI fragment that lacked the 5′ 1,005 bp of the BALF2 open reading frame from the cosmid cM966-20 (56) into the *Bgl*II-*Eco*RI-cleaved SCMV Colburn IE94 expression vector pGH179 to generate pEF55. The 5′ end of BALF2 was amplified by a polymerase chain reaction utilizing the following primers: at the 5′ end, 5′-CTAGGGATCCATGCAGGGTGCACAGACT-3′, and at the internal *Bgl*II site, 5′-GCAAAGATCTGCGTGGACAC-3′. The polymerase chain reaction mixtures contained 10 μl of 10× reaction buffer (Cetus), 10.0 μl of deoxynucleoside triphosphates (1.25 μM each dATP, dCTP, dGTP, and TTP), 5 μl of each primer at 20 μM, 10 ng of the appropriate plasmid DNA, and 0.5 μl of *Taq* polymerase (Cetus). The reaction mixtures were incubated in a thermocycler for 2 min at 94°C, 2 min at 52°C, and 3 min at 72°C for 40 cycles. Aliquots of the reaction mixtures were analyzed on an agarose gel for the presence of the desired 1,005-bp fragment, after which the DNA was digested with *Bam*HI and

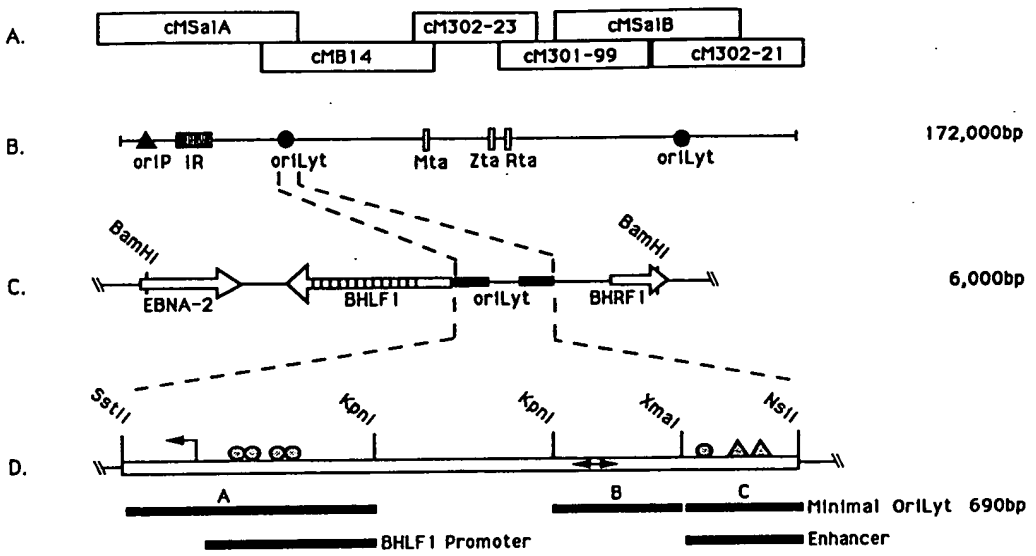

FIG. 1. Diagrammatic representation of the EBV genome showing the map locations of the genes of interest and the structure of ori-Lyt. (A) Six overlapping cosmids from the EBV strain M-ABA that were utilized to provide the entire EBV genome. (B) EBV genome. The locations of the *Bam*HI-W internal repeat (striped rectangle) and the origins of replication, ori-P (triangle) and ori-Lyt (circles), as well as the locations of the three lytic-cycle transactivators, Zta, Rta, and Mta (open rectangles), are shown. (C) 6-kb *Bam*HI-H fragment from EBV that served as the target in the replication assays. *Bam*HI-H contains 3′ sequences of the EBNA-2 gene, the entire BHLF1 open reading frame, ori-Lyt, and 5′ sequences of the BHRF1 gene. (D) Minimal ori-Lyt, which is contained within a 690-bp *Sst*II-*Nsi*I fragment, is composed of three domains. Domain A contains the promoter and leader region of the BHLF1 gene and four Zta-binding sites (ZREs, shaded circles). Domain B contains two AT-rich palindromes, one 18 bp and the other 20 bp, designated by the double-headed arrow. Domain C is composed of an enhancer element, which contains one ZRE (shaded circle) and two binding sites for the Rta transactivator (shaded triangles). The central *Kpn*I fragment is not required for ori-Lyt function (32).

*Bgl*II and cloned into pEF55 at the *Bgl*II site. The correct orientation regenerated the intact BALF2 open reading frame (pEF56A). The B95-8 DNA polymerase expression plasmid was a gift from Don Coen (Harvard University) and contains the DNA polymerase open reading frame BALF5 under the control of the simian virus 40 early promoter. The *Sal*I-F subclone (pGD4) contains the *Sal*I-F fragment from B95-8 cloned into pBR322 at the *Sal*I site. The *Bam*HI-BG clone (pDH33) contains the *Bam*HI-B and *Bam*HI-G fragments from P3HR-1 cloned into pBR322 at the *Bam*HI site. The plasmid pJMH4 (*Bam*HIΔBΔG) is an M-ABA subclone containing the 3′ 6,107 bp of *Bam*HI-B and the 5′ 3,003 bp of *Bam*HI-G cloned into the pHC79 vector.

RESULTS

Establishment of an ori-Lyt replication assay. To identify viral proteins required to replicate ori-Lyt, we established a transient cotransfection assay similar to that described by Challberg (8) for HSV type 1 (HSV-1). In that assay, large restriction fragments of the HSV-1 genome were cotransfected into Vero cells with a plasmid containing the HSV-1 origin of replication, ori-S, and *Dpn*I sensitivity was used to discriminate between input and replicated DNA (55). The *Dpn*I assay is based on the differential ability of *Dpn*I to cleave the input bacterially synthesized DNA which is methylated on the A residue of the *Dpn*I cleavage site GATC and the lack of cleavage when this methylation is lost after replication in eukaryotic cells. In the experiments described here, Vero cells (an African green monkey kidney cell line) were also used, and the exogenous EBV genome was provided by a set of six overlapping cosmids from the EBV strain M-ABA (56) (Fig. 1). The replication origin was provided by a plasmid carrying the EBV *Bam*HI-H fragment which encompasses the DS-L ori-Lyt. The transfection mixture also contained expression plasmids for the lytic-cycle transactivators, Zta, Rta, and Mta, to ensure adequate expression of the EBV early genes from the transfected cosmid clones. To determine the optimal assay conditions, we transfected Vero cells with the complete set of cosmids, the *Bam*HI-H target, and increasing amounts (0 to 2.5 μg) of the expression plasmids encoding the transactivators. The isolated cellular DNA was cleaved with restriction enzymes, and the DNA fragments were separated by gel electrophoresis and transferred onto a nylon membrane. The input, transfected DNA was visualized by hybridization with a radiolabeled pBR322 probe. The *Bam*HI-cleaved DNA is shown in Fig. 2A, and *Bam*HI-plus-*Dpn*I-cleaved DNA is shown in Fig. 2B. At a concentration of 2.5 μg of each expression plasmid, maximal replication of the target occurred (lane 4), as indicated by the presence of the *Dpn*I-resistant ori-Lyt band (arrowed). If no transactivators were included, no detectable replication occurred (lane 1). Below 2.5 μg of the transactivators, either no detectable replication occurred (lane 2) or minimal replication of the target occurred (lane 3). Replication was origin specific. Of all the input DNAs, only the *Bam*HI-H target (and to a lesser extent the cosmids that contain DS-R and DS-L) showed *Dpn*I resistance.

Replication genes are located on three separate cosmids. Having established the assay conditions, we next determined whether any of the cosmids were dispensable. The assay was repeated with one of the cosmids being omitted from each transfection mix (Fig. 3B). A *Dpn*I-resistant band was observed in the absence of cMSalB, cM302-23, or cMSalA, indicating that these three cosmids were not re-

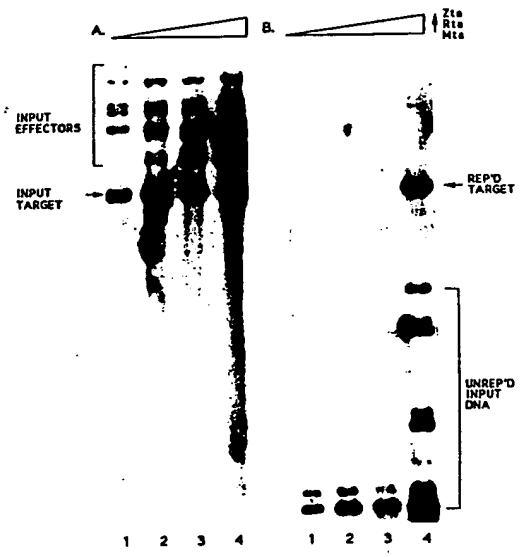

FIG. 2. Six overlapping cosmids support replication of the *Bam*HI-H target in the presence of the lytic-cycle transactivators. (A) Southern blot of transfected cell DNA cut with *Bam*HI and probed with radiolabeled pBR322 DNA. The top four bands of input DNA represent cosmids, and the central three bands represent the Zta, Rta, and Mta transactivators. Lane 1, no transactivators added; lanes 2 to 4, increasing amounts (0.1, 0.5, and 2.5 μg, respectively) of the transactivator expression plasmids. (B) As in panel A, except that the DNA was digested with *Bam*HI plus *Dpn*I. Replicated *Dpn*I-resistant target is readily detected with 2.5 μg of the Zta, Rta, and Mta expression constructs (lane 4).

quired for replication. On the other hand, if cM302-21, cM301-99, or cMB14 was removed, replication was not detected. Thus, in the presence of the lytic-cycle transactivators, only three of the cosmids, cMB14, cM301-99, and cM302-21, were required for replication of ori-Lyt.

To establish that the three essential cosmids, cMB14, cM301-99, and cM302-21, were sufficient for replication of

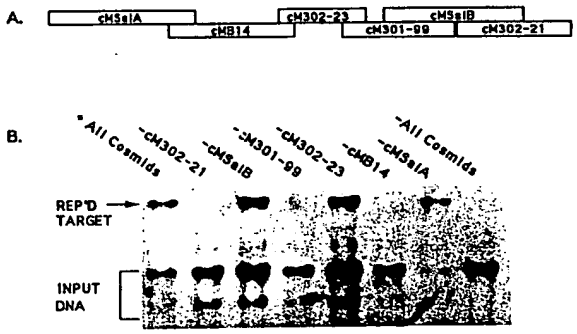

FIG. 3. Three cosmids, cMB14, cM301-99, and cM302-21, are each required for replication of the target. (A) Set of overlapping cosmids. (B) Southern blot of transfected cell DNA digested with *Bam*HI plus *Dpn*I and probed with pBR322 to detect *Dpn*I-resistant, replicated DNA. The positive control contained all six cosmids transfected with 2.5 μg of the expression constructs for Zta, Rta, and Mta. The target was resistant to cleavage with *Dpn*I. Replication of the target was negative when the cosmid cM302-21, cM301-99, or cMB14 was omitted. Replication of the target was positive in the absence of the cosmid cMSalB, cM302-23, or cMSalA. In the negative control, all cosmids were omitted.

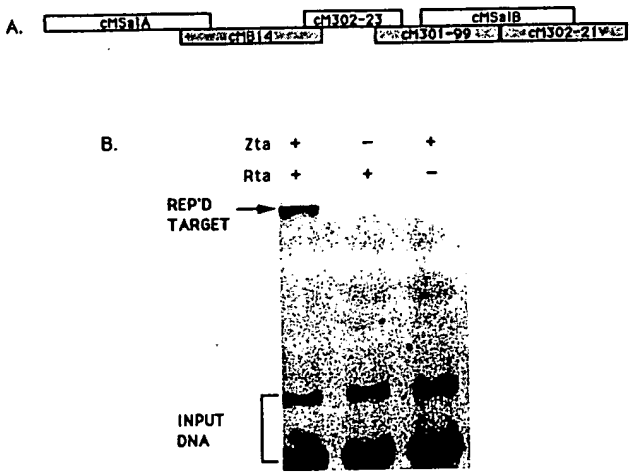

FIG. 4. Three essential cosmids, cMB14, cM301-99, and cM302-21, are sufficient for replication of the target, but only in the presence of the Zta and Rta expression plasmids. (A) Set of overlapping cosmids with the three essential cosmids highlighted. (B) Transfected cell DNA cut with *Bam*HI and *Dpn*I, Southern blotted, and probed with *Bam*HI-H to detect *Dpn*I-resistant, replicated DNA. The requirement for Zta and Rta was examined in cells cotransfected with cMB14, cM301-99, cM302-21, and Mta.

ori-Lyt, we cotransfected them with expression plasmids for Zta, Rta, and Mta. Efficient replication of the target occurred in the presence of these three cosmids (Fig. 4), demonstrating that cMB14, cM301-99, and cM302-21 not only were required for replication, but were sufficient. Addition of both the Zta and Rta transactivators remained a requirement for replication. If either was left out of the transfection mix, detectable replication of *Bam*HI-H did not occur. Thus, to obtain replication of ori-Lyt, three of the cosmids were required as well as two transcriptional activators, Zta and Rta. Zta and Rta could be functioning directly to transactivate the ori-Lyt promoter and enhancer or indirectly to increase expression from the cosmid-encoded genes needed for replication. The requirement for Mta could not be assessed in these experiments because Mta is encoded within one of the essential cosmids, cMB14.

ori-Lyt replication requires the EBV genes BALF5, BALF2, BMRF1, and BSLF1. The seven genes of HSV-1 whose products are essential and sufficient to replicate an HSV-1 origin-containing target plasmid are the DNA polymerase, UL30; the single-stranded DNA-binding protein, UL29; the tripartite helicase-primase complex containing UL5, UL8,

TABLE 1. Putative EBV homologs of the essential HSV replication proteins

HSV gene	EBV ORF[a]	Identity (%)	Function
UL30	BALF5	33	DNA polymerase (POL)
UL42	BMRF1	—[b]	Polymerase processivity factor (PPF)
UL29	BALF2	25	Single-stranded DNA-binding protein (SSB)
UL5	BBLF4	34	Helicase (HEL) } Complex
UL52	BSLF1	23	Primase (PRI) } Complex
UL8	BBLF2/3	—	Primase-associated factor (PAF)
UL9	?		Origin-binding protein (OBP)

[a] ORF, open reading frame.
[b] —, equivalent genomic location only.

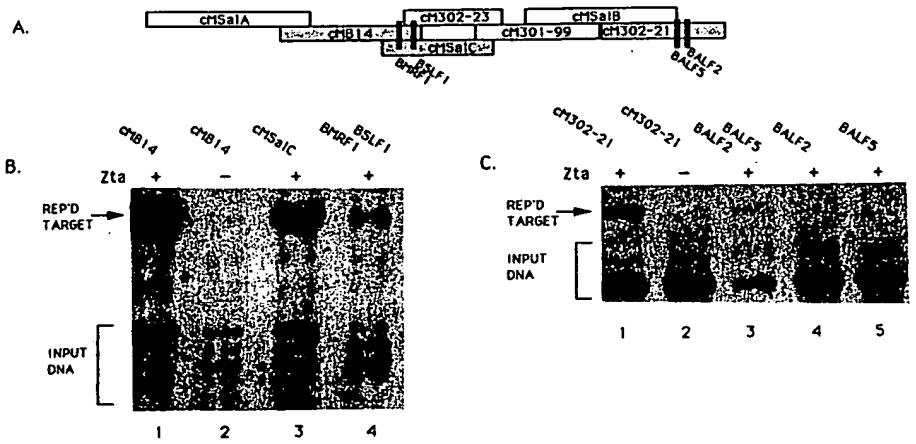

FIG. 5. Cosmids cMB14 and cM302-21 each encode only two essential replication genes. (A) Set of cosmids; cMB14, cMSalC, and cM302-21 are highlighted. The locations of the genes on cMB14, cMSalC, and cM302-21 that have homology with HSV-1 replication genes are also shown. In panels B and C, transfected cell DNA was cut with *Bam*HI and *Dpn*I, Southern blotted, and probed with *Bam*HI-H to detect *Dpn*I-resistant, replicated DNA. (B) Lane 1, the three cosmids, cMB14, cM301-99, and cM302-21, plus the transactivator expression plasmids. Lane 2, no Zta added. Lane 3, the cosmid cMB14 was replaced by the cosmid cMSalC. Lane 4, the cosmid cMB14 was replaced by expression plasmids encoding BMRF1, the polymerase processivity factor, and BSLF1, the primase homolog. (C) Lane 1, the three cosmids, cMB14, cM301-99, and cM302-21, plus the transactivator expression plasmids. Lane 2, no Zta added. Lane 3, the cosmid cM302-21 was replaced by expression plasmids encoding BALF5, the DNA polymerase, and BALF2, the single-stranded DNA-binding protein homolog. Lane 4, minus BALF5. Lane 5, minus BALF2.

and UL52; the DNA polymerase processivity factor, UL42; and the origin-binding protein, UL9. Table 1 shows the seven HSV-1 replication genes, their potential EBV homologs, and the percent homology which they share (52).

Upon inspection of the map location of those EBV genes that had potential homology with the HSV-1 replication genes, it became apparent that they were all located within the three cosmids that were essential for replication in our assay. To determine whether these genes indeed encoded functional replication proteins, the EBV DNA polymerase, BALF5, and its processivity factor, BMRF1, as well as BSLF1, BBLF4, and BALF2 were placed under the control of the strong simian virus 40 early or CMV major immediate-early promoters in eukaryotic expression vectors and tested in substitution experiments. First, the cosmid cMB14 was replaced with another cosmid, cMSalC, which has five open reading frames in common with cMB14. This cosmid was able to substitute functionally for cMB14 (Fig. 5B, lane 3), indicating that the replication genes provided by cMB14 were located within the region that overlaps with cMSalC. The common sequences contain BMRF1 and BSLF1 plus BMLF1 (Mta), BMRF2 (a late gene), and BORF2 and BaRF1, which encode the large and small subunits of ribonucleotide reductase (3, 24). The cosmid cMB14 was next replaced by expression plasmids for the two genes, BSLF1 and BMRF1, that were potential homologs for HSV replication genes. Replication was positive in the presence of BSLF1 and BMRF1, although the replication signal was reduced compared with that obtained with cMB14 or cMSalC (Fig. 5B, lane 4). Transfection of the three essential cosmids in the presence of the lytic-cycle transactivators served as the positive control (Fig. 5B, lane 1). Removal of Zta from the transfection mix abolished detectable replication of the target and served as the negative control (Fig. 5B, lane 2). In a parallel experiment, the cosmid cM302-21 was successfully replaced by expression plasmids for the two predicted replication genes that mapped within cM302-21, the DNA polymerase, BALF5, and BALF2, the putative single-stranded DNA-binding protein (Fig. 5C, lane 3). If either BALF5 (Fig. 5C, lane 4) or BALF2 (Fig. 5C, lane 5) was omitted, replication of the target did not occur. Therefore, the cosmid cMB14 encodes only two essential replication proteins, BSLF1, the primase homolog, and BMRF1, the processivity factor. Similarly, the cosmid cM302-21 also encodes only two replication proteins, the DNA polymerase, BALF5, and BALF2, the single-stranded DNA-binding protein homolog.

ori-Lyt replication also requires BBLF4, BBLF2/3, and *Sal*I-F. A diagram of the remaining essential cosmid, cM301-99, showing the map locations of relevant open reading frames is presented in Fig. 6A. The only gene encoded by cM301-99 which has recognized sequence homology to a replication protein of HSV-1 is BBLF4. However, when this cosmid was replaced by an expression plasmid for BBLF4, no detectable replication occurred (data not shown). Replication of the target was restored, however, if cM301-99 was replaced by the BBLF4 expression plasmid and two cM301-99 subclones, *Sal*I-F and *Bam*HI-BG, neither of which contains an intact BBLF4 open reading frame (Fig. 6B, lane 3). Removal of either *Bam*HI-BG (Fig. 6B, lane 4) or *Sal*I-F (Fig. 6B, lane 5) abolished detectable replication of the target. The three cosmids plus transactivators again served as the positive control (Fig. 6B, lane 1), while the negative control lacked Zta (Fig. 6B, lane 2). Removal of BBLF4 also abolished detectable replication of the target (data not shown). Because two cM301-99 subclones and the BBLF4 expression plasmid were required to replace cM301-99, at least three replication genes appear to be provided by this cosmid.

To determine the gene or genes provided by *Bam*HI-BG, we used subclones of this DNA fragment. The clone *Bam*HIΔBΔG(pJMH4) was able to substitute for *Bam*HI-BG in the replication assay (Fig. 6C, lane 4). Both *Bam*HI-BG and *Bam*HIΔBΔG contain the two open reading frames BBLF2 and BBLF3 (Fig. 6A), which are potential UL9 and UL8 homologs based on their genome locations (52). The linkage with UL8 was further strengthened by the recognition that BBLF2 has a stretch of 55 amino acids which are

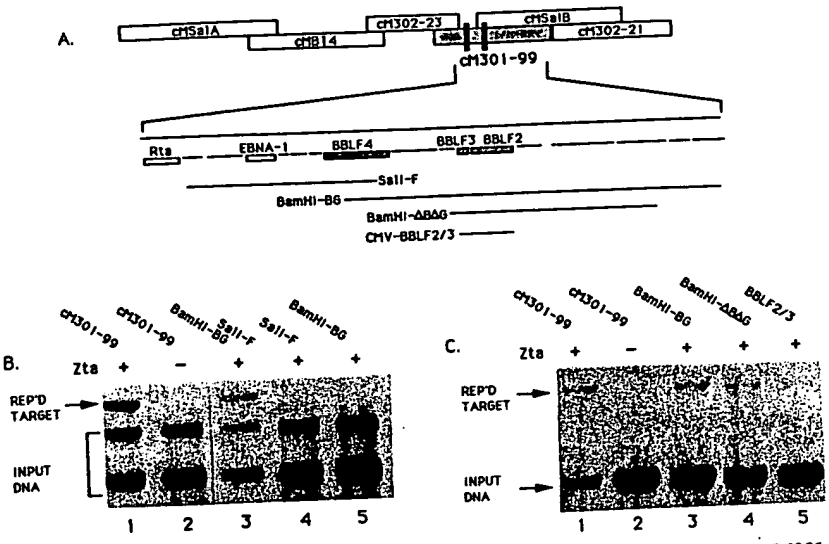

FIG. 6. Cosmid cM301-99 encodes at least three replication genes. (A) Set of overlapping cosmids; cM301-99 is highlighted, and the locations of the genes on cM301-99 known to be required for replication are designated. In the expanded segment of cM301-99, the structure of subclones of this region is shown, along with the location of relevant open reading frames. In panels B and C, transfected cell DNA was cut with *Bam*HI and *Dpn*I, Southern blotted, and probed with *Bam*HI-H (B) or with a subfragment of *Bam*HI-H (C) to detect *Dpn*I-resistant, replicated DNA. (B) Lane 1, the three essential cosmids plus the transactivator expression plasmids. Lane 2, no Zta added. Lane 3, the cosmid cM301-99 was replaced with the BBLF4 expression plasmid and the two subclones, *Sal*I-F and *Bam*HI-BG plus the transactivator expression plasmids. Lane 4, no *Bam*HI-BG added. Lane 5, no *Sal*I-F added. (C) Lane 1, the three essential cosmids transfected with the transactivator expression plasmids. Lane 2, no Zta added. Lane 3, cells were transfected with cMB14, cM302-21, the BBLF4 expression plasmid, the two subclones, *Sal*I-F and *Bam*HI-BG, as well as the transactivator expression plasmids. Lane 4, *Bam*HI-BG was replaced by *Bam*HIΔBΔG. Lane 5, *Bam*HI-BG was replaced by a CMV-driven expression plasmid encoding BBLF2 and BBLF3.

similar to UL8 of HSV-1. This region of similarity is also found in the VZV gene 52 and the HCMV gene UL102 (Fig. 7). Because of this sequence and positional homology, we cloned these two open reading frames into an expression vector. This BBLF2/3 expression construct, when cotransfected with *Sal*I-F and the BBLF4 expression plasmid, successfully substituted for *Bam*HIΔBΔG (Fig. 6C, lane 5). We conclude that the cosmid cM301-99 encodes at least three replication proteins. One of these is the helicase homolog encoded by the open reading frame BBLF4. The open reading frames BBLF2 and BBLF3 are thought to be spliced into one message in EBV (24) and hence would encode only one protein, potentially functioning as the UL8 homolog. The other gene (or genes) required for replication of ori-Lyt is located on the subclone *Sal*I-F.

The cotransfection assays identified six genes encoded by EBV whose products were individually essential for transient replication of an ori-Lyt-containing target in Vero cells. We also demonstrated that the combination of these six genes plus *Sal*I-F and the Zta, Rta, and Mta transcriptional transactivators were sufficient to support replication of ori-Lyt (Fig. 8, lane 4). In an earlier experiment (Fig. 4), we demonstrated that Rta and Zta were both required for replication. To address the question of the requirement for the Mta transactivator, we used the set of six cloned genes plus *Sal*I-F. As shown in Fig. 8 (lane 5), replication of ori-Lyt was significantly reduced but not eliminated in the absence of added Mta, even though both Zta and Rta were present. Therefore, all three transactivators appear to be necessary for maximal replication.

Finally, our standard ori-Lyt target was the *Bam*HI-H fragment, which contains an intact BHLF1 gene in addition to ori-Lyt (Fig. 1). Since a plasmid with this open reading frame was present in all the replication assays shown, the potential existed for the BHLF1 gene product to be contributing to ori-Lyt replication. To address this question, we substituted a modified *Bam*HI-H target (pPL2A) carrying a 1,375-bp deletion in the BHLF1 open reading frame for *Bam*HI-H. This modified ori-Lyt target was still replication competent when cotransfected with the six cloned replication genes, plus *Sal*I-F and the three transactivators, indicating that BHLF1 is not essential (Fig. 8, lane 1).

DISCUSSION

The identity of the virally encoded *trans*-acting proteins involved in HSV DNA replication was established through a combination of genetic studies and the use of transient cotransfection-replication assays. Several of the HSV replication genes have potential homologs in EBV based on sequence similarities. However, with the exception of the EBV DNA polymerase (BALF5) and the associated polymerase processivity factor (BMRF1), the functional equivalence of these genes has not previously been examined, nor has the full complement of genes required for lytic EBV DNA replication been determined. We established a cotransfection-replication assay utilizing Vero cells and overlapping cosmid clones of the EBV genome to examine the requirements for replication of an ori-Lyt-containing target plasmid. Our studies identified six EBV genes, BALF5, BMRF1, BALF2, BBLF4, BSLF1, and BBLF2/3, along with the viral lytic-cycle transactivators, Zta, Rta, and Mta, and an unidentified gene in *Sal*I-F, as being essential for ori-Lyt replication in the transient assay.

Two of the essential genes, the DNA polymerase and the polymerase processivity factor, were already known to serve a replication function. The first evidence for an EBV-encoded polymerase came from studies demonstrating that

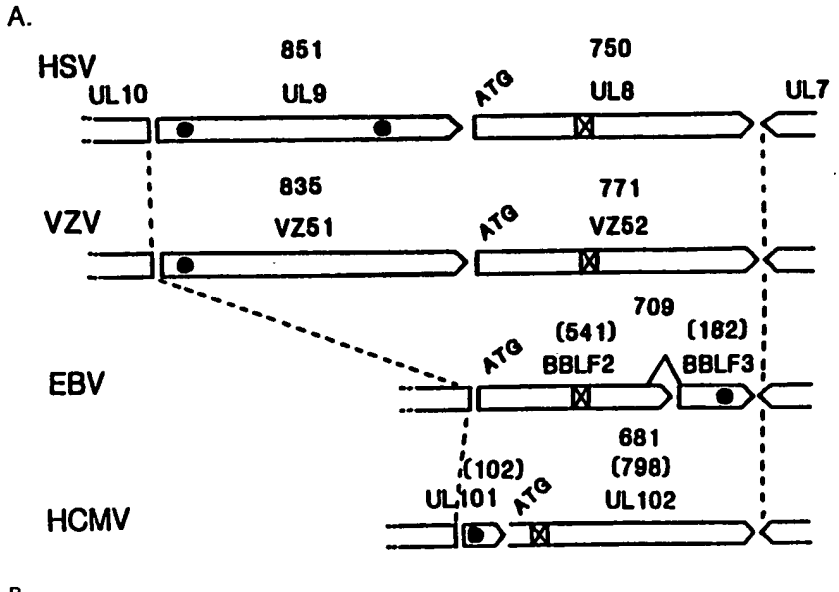

FIG. 7. BBLF2/3 shares genomic location and limited amino acid sequence similarity with UL8 of HSV, gene 51 of VZV, and UL102 of HCMV. (A) Relationship of the BBLF2 and BBLF3 open reading frames to the open reading frames in the equivalent regions of the HSV, VZV, and HCMV genomes (11, 19, 53). Potential ATP-binding motifs are indicated (●), as is a region of amino acid similarity (☒). (B) Comparison of the sequences denoted in panel A by the symbol ☒. Identical residues are indicated by asterisks (*), and equivalent residues are indicated by vertical lines (|).

viral lytic-cycle replication was selectively inhibited by drugs such as phosphonoacetic acid and acyclovir (16, 45). Sequencing of the EBV genome revealed an open reading frame (BALF5) whose protein product shares 33% sequence identity with the HSV DNA polymerase (3, 52) and like the HSV polymerase has several regions that are highly conserved in prokaryotic and eukaryotic viral DNA polymerases and in mammalian DNA polymerase alpha (64). More recently, BALF5 has been synthesized in vitro and shown to exhibit polymerase activity (50). The BMRF1 gene is the positional equivalent of the HSV UL42 polymerase processivity factor (20, 27, 29, 36). Although the EBV protein has no significant sequence similarity to UL42, it is a functional homolog. The BMRF1 protein copurifies with the EBV polymerase and has been shown to stimulate polymerase activity in vitro (40, 46).

The four other EBV genes identified as being required for ori-Lyt replication in the cotransfection-replication assay are BALF2, BBLF4, BSLF1, and BBLF2/3. The BALF2 gene product has 25% overall identity with the HSV single-stranded DNA-binding protein (UL29). BALF2 also contains a series of motifs that are highly conserved in the HSV, VZV, SCMV, and HCMV single-stranded DNA-binding proteins (1, 10, 65). The HSV and EBV proteins are virtually identical in size at 1,196 and 1,128 amino acids, respectively. The BBLF4 and BSLF1 proteins have significant sequence identity with the HSV UL5 and UL52 genes. UL5 and UL52 form a tripartite complex with UL8 in HSV-infected cells and in insect cells coinfected with recombinant baculoviruses expressing these genes (15, 21). This complex has both helicase and primase activities. Recently, helicase and pri-

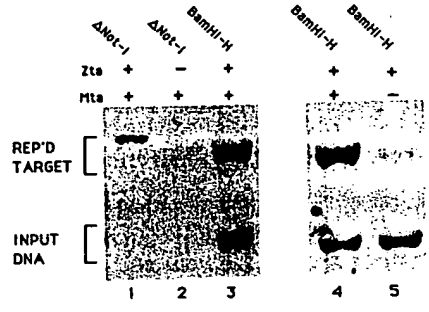

FIG. 8. Six cloned replication genes plus *Sal*I-F and the lytic-cycle transactivators support replication of the target; the BHLF1 gene product is not required for replication. The ori-Lyt-containing target was transfected with expression constructions for the six cloned genes: BMRF1, the polymerase processivity factor; BSLF1, the primase homolog; BBLF4, the helicase homolog; BBLF2/3, the potential UL8 homolog; BALF5, the DNA polymerase; and BALF2, the single-stranded DNA-binding protein homolog. *Sal*I-F was also added, as well as expression constructions for the three transactivators. A Southern blot of transfected cell DNA digested with *Bam*HI and *Dpn*I was probed with *Bam*HI-H to detect *Dpn*I-resistant, replicated DNA. Lane 1, an ori-Lyt target (ΔNotI) carrying a deletion within the BHLF1 open reading frame. Lane 2, no Zta added. Lane 3, the standard *Bam*HI-H target. Lane 4, the standard *Bam*HI-H target in the presence of Zta, Rta, and Mta. Lane 5, no Mta added.

mase activities were demonstrated in a bipartite complex of UL5 and UL52 (6, 22). Analysis of the UL5 amino acid sequence has revealed six motifs, including an ATP-binding motif, that are present in other families of helicases (28, 37, 44), and hence UL5 and BBLF4 are predicted to function as helicases. Indeed, individual mutations in each of the six conserved motifs of UL5 abolish its ability to complement a replication-deficient null mutant in a transient replication assay (76). UL52 is the putative primase of HSV, although an association with UL8 and UL5 may be required for complete activity. BSLF1 is the candidate primase of EBV. Furthermore, we believe that BBLF2/3 is the homolog of HSV UL8, the third member of the helicase-primase complex. BBLF3 and BBLF2 are the positional equivalents of HSV UL8 and UL9. However, in EBV, these two open reading frames are believed to be spliced into a single transcript and hence would presumably encode only one protein. Previous reports noted no significant similarity between BBLF2/3 and UL8 or UL9. However, a visual alignment (Fig. 7) reveals a 55-amino-acid region of BBLF2 that is conserved in HSV UL8, VZV 52, and HCMV UL102. For this reason, we believe that BBLF2/3 is likely to be the homolog of HSV UL8. The spliced BBLF2/3 transcript would encode 709 amino acids compared with 750 for HSV UL8.

At least one other gene encoded by EBV *Sal*I-F is required for ori-Lyt replication in the cotransfection assay. Interestingly, *Sal*I-F encodes the latency origin-binding protein EBNA-1. However, EBNA-1 is thought not to be the required gene product because *Sal*I-F cannot be replaced by other subclones that contain the EBNA-1 open reading frame (unpublished data). One of the replication proteins that has not been identified is an ori-Lyt origin-binding protein equivalent to the HSV origin-binding protein, UL9 (23, 41, 54). Since ori-Lyt is unrelated in sequence to HSV ori-S and ori-L, it is not surprising that comparative analyses have not revealed a homolog for UL9. It is possible that the gene for the EBV origin-binding protein is located in *Sal*I-F or that one of the proteins already shown to be required in the cotransfection assay is multifunctional and also provides origin-binding activity. Another alternative is that the origin-binding protein is not virally encoded and that a cellular factor serves in this capacity.

The EBV-encoded proteins identified in this study as being required for ori-Lyt replication are homologs of HSV proteins that participate directly in DNA replication. Herpesviruses also encode a number of enzymes, alkaline nuclease, ribonucleotide reductase, thymidine kinase, dUTPase, and uracil DNA glycosylase, that are involved in nucleotide metabolism and play an ancillary role in DNA synthesis (reviewed in reference 66). Genetic studies with HSV indicate that under certain conditions such as growth at high temperature, in growth-arrested cells, or in the animal host, HSV DNA replication may become dependent on the virally encoded alkaline nuclease, thymidine kinase, and ribonucleotide reductase. In our cotransfection-replication assays, the cosmid cMB14 could be replaced by a cosmid cMSalC which has five open reading frames in common with cMB14. The common genes are BMRF1 (the polymerase processivity factor), BSLF1 (the putative primase homolog), BMLF1 (the Mta transactivator), BMRF2, BORF2, and BaRF1. When these cosmids were replaced by expression vectors for BMRF1 and BSLF1, replication of ori-Lyt did occur but at a reduced level. This observation raises the possibility that cMB14 and cMSalC were providing an additional nonessential, replication-related function. Mta can be discounted since it was being provided exogenously by an expression vector, and BMRF2 is a late gene and therefore not a likely candidate for a replication-related function. The remaining common open reading frames, BORF2 and BaRF1, encode the large and small subunits of ribonucleotide reductase. One interpretation of the data is that the EBV-encoded ribonucleotide reductase is providing an auxiliary function that increases replication efficiency in these assays. The replication signal was also reduced when the *Bam*HI-H ori-Lyt target was replaced with a modified target carrying a deletion in the BHLF1 gene. In this case, the apparent reduction in replication efficiency may simply represent decreased hybridization of the probe to the target DNA which no longer contains the *Not*I repeats.

The minimal ori-Lyt as defined by Hammerschmidt and Sugden (32) consists of three essential subdomains (Fig. 1). Two of these, the BHLF1 promoter and the upstream enhancer, also function in transcriptional regulation of the flanking BHLF1 and BHRF1 genes. Characterized replication origins are commonly associated with transcriptional elements (reviewed in reference 19). These elements may be integrally linked to replication functions or may serve an auxiliary role by contributing to replication efficiency. For example, the cloned minimal HSV-1 origin, ori-S, can function in transient replication assays. However, if the flanking sequences containing the divergent promoters for the immediate-early IE175 and IE68 genes are also included, replication of the target is substantially increased (69). The EBV latency origin of replication (ori-P) consists of two domains, an EBNA-1-dependent enhancer and a region of dyad symmetry that is the site of initiation of replication (26, 72). With ori-P, replication is strictly dependent on the presence of the enhancer domain. The recently identified CMV origin of replication contains multiple binding sites for transcription factors (2, 33), although their requirement for replication has yet to be addressed by mutational analyses.

The BHLF1 promoter that constitutes one of the essential domains of EBV ori-Lyt contains four binding sites for Zta and is efficiently activated by Zta in cotransfection assays. The ori-Lyt enhancer contains two binding sites for a second viral transcriptional activator, Rta, and one binding site for Zta (31, 48). The enhancer is strongly activated by Rta and responds synergistically to the combination of Rta and Zta (14). In the cotransfection-replication assays, ori-Lyt replication was dependent on the presence of Zta, Rta, and Mta. In the assays using the cosmid clones, the transactivators would have been required for efficient expression of the replication genes encoded within the cosmids. Even in the final assays described here, at least one of the genes necessary for replication was provided on *Sal*I-F, and again it is likely that the transactivators would be needed for expression from this plasmid. In HSV, the immediate-early transactivators were required for replication of ori-S in the transient replication assay when viral DNA fragments were used to provide the replication functions. In contrast, when each of the seven replication genes was expressed from the strong constitutive HCMV promoter, the requirement for these transcription factors in the transient replication assay was alleviated (35). In EBV, however, the inclusion of the promoter and enhancer elements within the defined limits of the minimal ori-Lyt makes it highly probable that Zta and Rta contribute to replication directly through transcriptional activation, through DNA binding, or through interactions with components of the replication complex. The role of these transactivators in ori-Lyt replication can be addressed

more directly when the full complement of EBV replication genes has been identified.

ACKNOWLEDGMENTS

We thank Don Coen and J. Marie Hardwick for gifts of plasmids, Mabel Chiu for technical assistance, and Kimberley Collins, Sarah Heaggans, and Pamela Wright for assistance in preparation of the manuscript.

This work was funded by Public Health Service grants R37 CA30356 (S.D.H.) and R01 CA22130 (G.S.H). E.F. was supported by training grants 5T32 CA09243 and 5T32 NS07000.

REFERENCES

1. **Anders, D. G., and W. Gibson.** 1988. Location, transcript analysis, and partial nucleotide sequence of the cytomegalovirus gene encoding an early DNA-binding protein with similarities to ICP8 of herpes simplex virus type 1. J. Virol. 62:1364–1372.
2. **Anders, D. G., and S. M. Punturieri.** 1991. Multicomponent origin of cytomegalovirus lytic-phase DNA replication. J. Virol. 65:931–937.
3. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) 310:207–211.
4. **Baumann, R. P., R. R. Yalamanchili, and D. J. O'Callaghan.** 1989. Functional mapping and DNA sequence of an equine herpesvirus 1 origin of replication. J. Virol. 63:1275–1283.
5. **Buisson, M., E. Manet, M.-C. Trescol-Biemont, H. Gruffat, B. Durand, and A. Sergeant.** 1989. The Epstein-Barr virus (EBV) early protein EB2 is a posttranscriptional activator expressed under the control of EBV transcription factors EB1 and R. J. Virol. 63:5276–5284.
6. **Calder, J. M., and N. D. Stow.** 1990. Herpes simplex virus helicase-primase: the UL8 protein is not required for DNA-dependent ATPase and DNA helicase activities. Nucleic Acids Res. 18:3573–3578.
7. **Camp, H. S., P. M. Coussens, and R. F. Silva.** 1991. Cloning, sequencing, and functional analysis of a Marek's disease virus origin of DNA replication. J. Virol. 65:6320–6324.
8. **Challberg, M. D.** 1986. A method for identifying the viral genes required for herpesvirus DNA replication. Proc. Natl. Acad. Sci. USA 83:9094–9098.
9. **Chang, Y.-N., D. L.-Y. Dong, G. S. Hayward, and S. D. Hayward.** 1990. The Epstein-Barr virus Zta transactivator: a member of the bZIP family with unique DNA-binding specificity and a dimerization domain that lacks the characteristic heptad leucine zipper motif. J. Virol. 64:3358–3369.
10. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchinson III, T. Kouzarides, J. A. Martignetti, E. Preddie, S. C. Satchwell, P. Tomlinson, K. M. Weston, and B. G. Barrell.** 1990. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. 154:126–169.
11. **Chen, C., and H. Okayama.** 1987. High-efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. 7:2745–2752.
12. **Cho, M.-S., K.-T. Jeang, and S. D. Hayward.** 1985. Localization of the coding region for an Epstein-Barr virus early antigen and inducible expression of this 60-kilodalton nuclear protein in transfected fibroblast cell lines. J. Virol. 56:852–859.
13. **Countryman, J., and G. Miller.** 1985. Activation of expression of latent Epstein-Barr herpesvirus after gene transfer with a small cloned subfragment of heterogeneous viral DNA. Proc. Natl. Acad. Sci. USA 81:7632–7636.
14. **Cox, M. A., J. Leahy, and J. M. Hardwick.** 1990. An enhancer within the divergent promoter of Epstein-Barr virus responds synergistically to the R and Z transactivators. J. Virol. 64:313–321.
15. **Crute, J. J., T. Tsurumi, L. Zhu, S. K. Weller, P. D. Olivo, M. D. Challberg, E. S. Mocarski, and I. R. Lehman.** 1989. Herpes simplex virus 1 helicase-primase: a complex of three herpes-encoded gene products. Proc. Natl. Acad. Sci. USA 86:2186–2189.
16. **Datta, A. K., B. M. Colby, J. E. Shaw, and J. S. Pagano.** 1980. Acyclovir inhibition of Epstein-Barr virus replication. Proc. Natl. Acad. Sci. USA 77:5163–5166.
17. **Davison, A. J., and P. Taylor.** 1987. Genetic relations between varicella-zoster virus and Epstein-Barr virus. J. Gen. Virol. 68:1067–1079.
18. **Delius, H., and G. W. Bornkamm.** 1978. Heterogeneity of Epstein-Barr virus. III. Comparison of a transforming and a non-transforming virus by partial denaturation mapping of their DNAs. J. Virol. 27:81–89.
19. **DePamphilis, M. L.** 1988. Transcriptional elements as components of eukaryotic origins of DNA replication. Cell 52:635–638.
20. **Digard, P., and D. M. Coen.** 1990. A novel functional domain of an α-like DNA polymerase. J. Biol. Chem. 265:17393–17396.
21. **Dodson, M. S., J. J. Crute, R. C. Bruckner, and I. R. Lehman.** 1989. Overexpression and assembly of the herpes simplex virus type 1 helicase-primase in insect cells. J. Biol. Chem. 264: 20835–20838.
22. **Dodson, M. S., and I. R. Lehman.** 1991. Association of DNA helicase and primase activities with a subassembly of the herpes simplex virus 1 helicase-primase composed of the UL5 and UL52 gene products. Proc. Natl. Acad. Sci. USA 88:1105–1109.
23. **Elias, P., M. E. O'Donnell, E. S. Mocarski, and I. R. Lehman.** 1986. A DNA binding protein specific for an origin of replication of herpes simplex virus type 1. Proc. Natl. Acad. Sci. USA 83:6322–6326.
24. **Farrell, P. J.** 1989. Epstein-Barr virus genome. Adv. Viral Oncol. 8:103–132.
25. **Feinberg, A., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6–13.
26. **Gahn, T. A., and C. L. Schildkraut.** 1989. The Epstein-Barr virus origin of plasmid replication, oriP, contains both the initiation and termination sites of DNA replication. Cell 58:527–535.
27. **Gallo, M., D. Jackwood, M. Murphy, H. Marsden, and D. Parris.** 1988. Purification of the herpes simplex virus type 1 65-kilodalton DNA-binding protein: properties of the protein and evidence of its association with the virus-encoded DNA polymerase. J. Virol. 62:2874–2883.
28. **Gorbalenya, A. E., E. B. Koonin, Z. P. Donchenko, and V. M. Blinov.** 1988. A conserved NTP motif in putative helicases. Nature (London) 333:22.
29. **Gottlieb, J., A. I. Marcy, D. M. Coen, and M. D. Challberg.** 1990. The herpes simplex type 1 UL42 gene product: a subunit of DNA polymerase that functions to increase processivity. J. Virol. 64:5976–5987.
30. **Graham, F., and A. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52:456–467.
31. **Gruffat, H., E. Manet, A. Rigolet, and A. Sergeant.** 1990. The enhancer factor R of Epstein-Barr virus (EBV) is a sequence-specific DNA binding protein. Nucleic Acids Res. 18:6835–6843.
32. **Hammerschmidt, W., and B. Sugden.** 1988. Identification and characterization of oriLyt, a lytic origin of DNA replication of Epstein-Barr virus. Cell 55:427–433.
33. **Hamzeh, F. M., P. S. Lietman, W. Gibson, and G. S. Hayward.** 1990. Identification of the lytic origin of DNA replication in human cytomegalovirus by a novel approach utilizing ganciclovir-induced chain termination. J. Virol. 64:6148–6195.
34. **Hardwick, J. M., P. M. Lieberman, and S. D. Hayward.** 1988. A new Epstein-Barr virus transactivator, R, induces expression of a cytoplasmic early antigen. J. Virol. 62:2274–2284.
35. **Heilbronn, R., and H. zur Hausen.** 1989. A subset of herpes simplex virus replication genes induces DNA amplification within the host cell genome. J. Virol. 63:3683–3692.
36. **Hernandez, T. R., and I. R. Lehman.** 1990. Functional interaction between the herpes simplex-1 DNA polymerase and UL42 protein. J. Biol. Chem. 265:11227–11232.

37. **Hodgman, T.** 1988. A new superfamily of replicative proteins. Nature (London) 333:22–23.
38. **Hummel, M., and E. Kieff.** 1982. Epstein-Barr virus RNA. VIII. Viral RNA in permissively infected B95-8 cells. J. Virol. 43:262–272.
39. **Kenney, S., J. Kamine, E. Holley-Guthrie, E. C. Mar, J. C. Lin, D. Markovitz, and J. Pagano.** 1989. The Epstein-Barr virus immediate-early gene product, BMLF1, acts in *trans* by a posttranscriptional mechanism which is reporter gene dependent. J. Virol. 63:3870–3877.
40. **Kiehl, A., and D. I. Dorsky.** 1991. Cooperation of EBV DNA polymerase and EA-D(BMFR1) *in vitro*, and colocalization in nuclei of infected cells. Virology 184:330–340.
41. **Koff, A., and P. Tegtmeyer.** 1988. Characterization of major recognition sequences for a herpes simplex virus type 1 origin-binding protein. J. Virol. 62:4096–4103.
42. **Kowalski, D., and M. J. Eddy.** 1989. The DNA unwinding element: a novel, *cis*-acting component that facilitates opening of the *Escherichia coli* replication origin. EMBO J. 8:4335–4344.
43. **Kupershmidt, S., J. M. De Marchi, Z. Lu, and T. Ben-Porat.** 1991. Analysis of an origin of DNA replication located at the L terminus of the genome of pseudorabies virus. J. Virol. 65:6283–6291.
44. **Lane, D.** 1988. Enlarged family of putative helicases. Nature (London) 334:478.
45. **Leinbach, S., J. Reno, L. Lee, A. Isbell, and J. Boezi.** 1976. Mechanism of phosphonoacetate inhibition of herpesvirus-induced DNA polymerase. Biochemistry 15:426–430.
46. **Li, J.-S., B.-S. Zhou, G. E. Dutschman, S. P. Grill, R.-S. Tan, and Y.-C. Cheng.** 1987. Association of Epstein-Barr virus early antigen diffuse component and virus-specified DNA polymerase activity. J. Virol. 61:2947–2949.
47. **Lieberman, P. M., J. M. Hardwick, and S. D. Hayward.** 1989. Responsiveness of the Epstein-Barr virus *Not*I repeat promoter to the Z transactivator is mediated in a cell-type-specific manner by two independent signal regions. J. Virol. 63:3040–3050.
48. **Lieberman, P. M., J. M. Hardwick, J. Sample, G. S. Hayward, and S. D. Hayward.** 1990. The Zta transactivator involved in induction of lytic cycle gene expression in Epstein-Barr virus-infected lymphocytes binds to both AP-1 and ZRE sites in target promoter and enhancer regions. J. Virol. 64:1143–1155.
49. **Lieberman, P. M., P. O'Hare, G. S. Hayward, and S. D. Hayward.** 1986. Promiscuous trans-activation of gene expression by an Epstein-Barr virus-encoded early nuclear protein. J. Virol. 60:140–148.
50. **Lin, J.-C., N. D. Sista, F. Besencon, J. Kamine, and J. S. Pagano.** 1991. Identification and functional characterization of Epstein-Barr virus DNA polymerase by in vitro transcription-translation of a cloned gene. J. Virol. 65:2728–2731.
51. **Lockshon, D., and D. A. Galloway.** 1988. Sequence and structural requirements of a herpes simplex viral DNA replication origin. Mol. Cell. Biol. 8:4018–4027.
52. **McGeoch, D. J., M. A. Dalrymple, A. Dolan, D. McNab, L. J. Perry, P. Taylor, and M. D. Challberg.** 1988. Structures of herpes virus type 1 genes required for replication of virus DNA. J. Virol. 62:444–453.
53. **Miller, G., M. Rabson, and L. Heston.** 1984. Epstein-Barr virus with heterogeneous DNA disrupts latency. J. Virol. 50:174–182.
54. **Olivo, P. D., N. J. Nelson, and M. D. Challberg.** 1988. Herpes simplex virus DNA replication: the UL9 gene encodes an origin-binding protein. Proc. Natl. Acad. Sci. USA 85:5414–5418.
55. **Peden, K., J. Pipas, S. Pearson-White, and D. Nathans.** 1980. Isolation of mutants of an animal virus in bacteria. Science 208:1392–1396.
56. **Polack, A., G. Hartl, U. Zimber, U.-K. Freese, G. Laux, K. Takaki, B. Hohn, L. Gissmann, and G. W. Bornkamm.** 1984. A complete set of overlapping cosmid clones of M-ABA virus derived from nasopharyngeal carcinoma and its similarity to other Epstein-Barr virus isolates. Gene 27:279–288.
57. **Rawlins, D. R., G. Milman, S. D. Hayward, and G. S. Hayward.** 1985. Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA-1) to clustered sites in the plasmid maintenance region. Cell 42:859–868.
58. **Reisman, D., and B. Sugden.** 1986. *trans*-activation of an Epstein-Barr viral (EBV) transcriptional enhancer by the EBV nuclear antigen 1. Mol. Cell. Biol. 6:3838–3846.
59. **Sato, H., T. Takimoto, S. Tanaka, J. Tanaka, and N. Raab-Traub.** 1990. Concatameric replication of Epstein-Barr virus: structure of the termini in virus-producer and newly transformed cell lines. J. Virol. 64:5295–5300.
60. **Sixbey, J. W.** 1989. Epstein-Barr virus and epithelial cells. Adv. Viral Oncol. 8:187–202.
61. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.
62. **Stow, N., and C. McMonagle.** 1983. Characterization of the TR-S/IR-S origin of DNA replication of herpes simplex virus type 1. Virology 130:427–438.
63. **Sugden, B., and N. Warren.** 1989. A promoter of Epstein-Barr virus that can function during latent infection can be transactivated by EBNA-1, a viral protein required for viral DNA replication during latent infection. J. Virol. 63:2644–2649.
64. **Wang, T. S.-F.** 1991. Eukaryotic DNA polymerases. Annu. Rev. Biochem. 60:513–552.
65. **Wang, Y., and J. D. Hall.** 1990. Characterization of a major DNA-binding domain in the herpes simplex type 1 DNA-binding protein (ICP8). J. Virol. 64:2082–2089.
66. **Weller, S. K.** 1990. Genetic analysis of HSV genes required for genome replication, p. 105–135. *In* E. K. Wagner (ed.), Herpesvirus transcription and its regulation. CRC Press, Inc., Boca Raton, Fla.
67. **Weller, S. K., A. Spadaro, J. E. Schaffer, A. W. Murray, A. M. Maxam, and P. A. Schaffer.** 1985. Cloning, sequencing, and functional analysis of oriL, a herpes simplex type 1 origin of DNA synthesis. Mol. Cell. Biol. 5:930–942.
68. **Wells, R. D.** 1988. Unusual DNA structures. J. Biol. Chem. 263:1095–1098.
69. **Wong, S. W., and P. A. Schaffer.** 1991. Elements in the transcriptional regulatory region flanking herpes simplex virus type 1 oriS stimulate origin function. J. Virol. 65:2601–2611.
70. **Wu, C. A., N. J. Nelson, D. J. McGeoch, and M. D. Challberg.** 1988. Identification of herpes simplex virus type 1 genes required for origin-dependent DNA synthesis. J. Virol. 62:435–443.
71. **Wysokenski, D. A., and J. L. Yates.** 1989. Multiple EBNA1-binding sites are required to form an EBNA1-dependent enhancer and to activate a minimal replicative origin within *ori*P of Epstein-Barr virus. J. Virol. 63:2657–2666.
72. **Yates, J., N. Warren, D. Reisman, and B. Sugden.** 1984. A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. Proc. Natl. Acad. Sci. USA 81:3806–3810.
73. **Yates, J. L., and N. Guan.** 1991. Epstein-Barr virus-derived plasmids replicate only once per cell cycle and are not amplified after entry into cells. J. Virol. 65:483–488.
74. **Yates, J. L., N. Warren, and B. Sugden.** 1985. Stable replication of plasmids derived from Epstein-Barr virus in a variety of mammalian cells. Nature (London) 313:812–815.
75. **Zehnbauer, B. A., D. Small, G. M. Brodeur, and B. Vogelstein.** 1988. Characterization of N-*myc* amplification units in human neuroblastoma cells. Mol. Cell. Biol. 8:522–530.
76. **Zhu, L., and S. K. Weller.** 1992. The six conserved helicase motifs of the UL5 gene product, a component of the herpes simplex virus type 1 helicase-primase, are essential for its function. J. Virol. 66:469–479.
77. **zur Hausen, H., F. O'Neill, and U. Freese.** 1978. Persisting oncogenic herpesvirus induced by the tumor promoter TPA. Nature (London) 272:373–375.